THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930-1938
EDITED BY S. S. WILKS, 1938-1949

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. 27, No. 1 — March, 1956

THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

INSTITUTE OF MATHEMATICAL STATISTICS

Treasurer's Office — Department of Statistics, Stanford University, Stanford, Cal.
A. H. Bowker, Treasurer
These addresses should be used for communications concerning subscriptions, back numbers, payment of dues, etc.

Secretary's Office — Department of Statistics, University of North Carolina, Chapel Hill, N. C.
George E. Nicholson, Jr., Secretary
These addresses should be used for communications concerning membership, changes of address, etc. Changes of address which are to become effective for a given issue of the *Annals* should be reported to the Secretary on or before the 10th of the month preceding the month of issue.

Editorial Office — The RAND Corporation, 1700 Main St., Santa Monica, Cal.
T. E. Harris, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Dittoed and mimeographed papers are acceptable if they are completely legible and printed on paper that will take corrections. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulas in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae. Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Prices — Current issues are $12 per volume (four issues of one calendar year) in the U. S. and Canada; $10 per volume elsewhere. Single issues are $3.50. Back numbers are $12 per volume, $3.50 per issue, $200 for the first 25 volumes, $10 per additional volume purchased at the same time as Volume I through XXV.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

NEW POLICY FOR ABSTRACTS

The Institute of Mathematical Statistics has a new policy on abstracts. They should be submitted in duplicate to the Editor, preferably on abstract blanks, which can be obtained from the IMS Secretary. Abstracts must be received at least 47 days before the first day of the meeting at which they are to be presented, indicating whether presented by title or in person. (Only one contributed paper may be given in person at any one meeting.) They may be printed prior to the publication of the report of the meeting. Those received by April 30 will appear in the September *Annals*, by July 31 in December, etc. Abstracts should be limited to 200 words or the equivalent, and should avoid displayed expressions and complicated formulae. They can be accepted from non-members of the IMS only if transmitted by members.

LARGE-SAMPLE THEORY: PARAMETRIC CASE[1]

By Herman Chernoff

Stanford University

1. Introduction. Large-sample theory is a branch of statistics which seems to have developed because the existence of certain theorems in the theory of probability made it relatively easy to obtain good approximate results if the sample size is large. These theorems, like the law of large numbers and the central limit theorem, are extremely elegant, and frequently their elegance is captured by these "easily" obtained results. This elegance has undoubtedly stimulated a great many people to do work in statistics.

However, since one is seldom faced with an infinite sample, it is relevant to ask whether asymptotic results are useful, and if so, where. In particular, one is often asked whether a given sample size is large enough to justify the use of asymptotic results. Frequently this question is embarrassing, and no answer is available simply because the answer would involve the solution of the more difficult finite-sample-size problem and the use of nonexistent related tables. In some cases, where this question has been treated, it has been shown that these asymptotic results are very good approximations. One example is the study wherein it was shown that the chi-square goodness-of-fit statistic has approximately the chi-square distribution for rather small sample sizes [1].

Even though results of this type are not available for a particular problem, the study of the large-sample case could be justified on other grounds. Asymptotic solutions of a problem frequently give insight into what constitutes a reasonable procedure for the finite-sample-size case. Everyone who has had the experience of seeing how obvious the solution to a certain problem is after spending hours deriving it can appreciate how suggestive an asymptotic result can be for the finite-sample-size problem. For somewhat similar reasons, the method of maximum likelihood estimation, which has various good large-sample properties, has become extremely popular, even for small samples. In fact, a glance at the literature gives the impression that the property of being a maximum likelihood estimate has almost been adopted as the criterion of optimality.

In this paper we deal with the parametric case. Ordinarily this is assumed to mean that our observations come from a population whose distribution is specified by the value of a parameter θ, which may be a k-dimensional vector. A specific problem would be that of testing whether two normal populations with the same variance have the same mean. It seems that once more we must face the fact that our problems may not reflect reality completely. There is a considerable class of problems for which the parametric formulation is more than a

[1] Presented as a special invited address at the Annual Meeting of the IMS in Berkeley, California, December 27, 1954. This work was prepared with the partial support of the Office of Naval Research.

convenient and very rough approximation. On the other hand, there is a considerable class for which this is not so. Even in these cases the same sort of reasoning which was advanced to advocate the study of large samples results is *apropos* to justify the study of parametric theory, and even of its application to problems where the parametric formulation seems quite rough.

Another point of some interest is that the normal distribution, on occasion, plays the role of a worst distribution. In such cases one may obtain quasi-maximum likelihood estimates, i.e., estimates derived by the use of maximum likelihood on the not necessarily correct assumption that certain random variables are normally distributed. These estimates may be inefficient compared with the true maximum likelihood estimates. Still, these quasi-maximum likelihood estimates have the same or as good asymptotic distributions as they would have were the assumptions of normality correct. They also have the advantage that their computation does not involve the knowledge of the true distribution of these variables. Some complex examples are treated in [2]. A trivial example which illustrates this point is the following. On the basis of a sample of n independent observations, estimate the mean of the population when it is assumed to be normal and it really is rectangular. Here, $\bar{X}$ is the quasi-maximum likelihood estimate and the true maximum likelihood estimate is $\hat{\mu}$, the average of the smallest and largest observations. The asymptotic distribution of $\sqrt{n}(\bar{X} - \mu)$ is normal with mean 0 and variance σ^2 (the variance of the population), whether the population is normal or rectangular. However, if it is rectangular, $\hat{\mu}$ will be considerably more efficient.

This paper will be divided into two main parts. In the first I shall summarize several techniques and results which are useful tools in the study of large-sample theory and which, I feel, have been unfortunately neglected in the literature. In the second part I shall consider some results in inference in the large-sample parametric case. There, much of the space will be devoted to material which has been of special interest to me. In this way I hope to communicate some of my outlook rather than merely to present a long list of accomplishments.

Part I

2. Stochastic limit and order relationships. The title of this section is taken from that of a paper of Mann and Wald [3]. Their stated purpose was to provide readers with certain general results which would eliminate the necessity on the part of future authors of laboriously proving special cases, not to mention confusing the readers. This aim seems to have been largely frustrated mainly by the fact that the paper was practically forgotten. I wish to discuss some of these general results and notations and some useful generalizations of these.

In standard notation one writes $a_n = O(r_n)$ if $\{a_n\}$ is a sequence of real numbers and $\{r_n\}$ is a sequence of positive numbers such that a_n / r_n is bounded. If $a_n / r_n \to 0$ as $n \to \infty$, one writes $a_n = o(r_n)$. This notation is frequently convenient and suggestive. For example, if $a_n \to 0$ and b_n is bounded, it follows that $a_n b_n \to 0$. This may be simply written as follows: $o(1)\, O(1) = o(1)$.

An analogous notation may be defined for sequences of chance variables $\{x_n\}$. We may write $x_n = O_p(r_n)$ (x_n / r_n is bounded in probability) if for each $\epsilon > 0$ there is an M_ϵ and an N_ϵ such that

$$\Pr\{|x_n| \geqq M_\epsilon r_n\} \leqq \epsilon \qquad \text{for } n \geqq N_\epsilon .$$

Finally, we may write $x_n = o_p(r_n)$ ($|x_n| / r_n$ approaches zero in probability) if

$$\Pr\{|x_n| \geqq \epsilon r_n\} \to 0 \qquad \text{as } n \to \infty \qquad \text{for each } \epsilon > 0.$$

It might be well to note here that these concepts are easily extendable to the case where x_n is not necessarily a real chance variable, but where x_n may take on values in an arbitrary space on which an "absolute value" is defined.

One of the results obtained by Mann and Wald is part of their Corollary 1, which states essentially that the algebra of o and O extends to o_p and O_p. A paraphrase of this result, which I have found very useful, is due to John Pratt and is stated as follows: Suppose that $\{x_n\}$ is a sequence of chance variables defined on an arbitrary space. Let $\{g_n(x_n)\}$ and $\{f_n^{(j)}(x_n)\}$, $j = 1, 2, \cdots, k$, be $k + 1$ sequences of measurable functions, and let $\{r_n\}$ and $\{r_n^{(j)}\}$ be $k + 1$ sequences of positive numbers.

THEOREM 1. *Suppose that*

$$(1) \qquad \begin{aligned} f_n^{(j)}(x_n) &= O_p(r_n^{(j)}), & j &= 1, 2, \cdots, k_1, \\ f_n^{(j)}(x_n) &= o_p(r_n^{(j)}), & j &= k_1 + 1, \quad k_1 + 2, \cdots, k, \end{aligned}$$

and that

(2) *for any (nonrandom) sequence* $\{a_n\}$ *for which*

$$f_n^{(j)}(a_n) = O(r_n^{(j)}), \qquad j = 1, 2, \cdots, k_1,$$

and

$$f_n^{(j)}(a_n) = o(r_n^{(j)}), \quad j = k_1 + 1, \quad k_1 + 2, \cdots, k,$$

hold, it follows that $g_n(a_n) = O(r_n)$.

Then, it follows that $g_n(x_n) = O_p(r_n)$. *Furthermore, if the last line of* (2) *is replaced by* $g_n(a_n) = o(r_n)$, *the conclusion is* $g_n(x_n) = o_p(r_n)$.

The following are some examples which may serve to illustrate the use of this result.

EXAMPLE 1. If $y_n \underset{p}{\to} y$, i.e., if y_n approaches y in probability, or $y_n - y = o_p(1)$, and if $z_n \underset{p}{\to} z$, then $y_n z_n \underset{p}{\to} yz$. This result follows because we are given, on the one hand, that $y_n - y = o_p(1)$ and $z_n - z = o_p(1)$. On the other hand, it is easy to prove (and is well known) that $b_n - b = o(1)$ and $c_n - c = o(1)$ (i.e., that $b_n \to b$ and $c_n \to c$) imply that $b_n c_n - bc = o(1)$. Consequently,

$$y_n z_n - yz = o_p(1).$$

Several remarks may be made about this example. It may seem to involve a tremendous amount of machinery for a very simple result. In fact, a direct proof

may seem to be no more difficult than the "on-the-other-hand" part. Actually, my own experience in class has shown that the direct proof is usually instructive because students find it so difficult. The tremendous machinery is not so tremendous if this approach is used frequently, for then it becomes standard. Finally, this example illustrates how this approach clearly separates the nonstochastic asymptotic elements of a problem from the stochastic elements.

One point which may have not been thoroughly clarified in the above exposition is the specification of x_n, $f_n^{(j)}$, g_n, and a_n in this example. To be perfectly specific, we may let

$$x_n = (y_n, z_n, y, z); \qquad f_n^{(1)}(x_n) = y_n - y;$$
$$g(x_n) = y_n z_n - yz; \qquad f_n^{(2)}(x_n) = z_n - z;$$
$$a_n = (b_n, c_n, b, c).$$

EXAMPLE 2. If $x_n = o_p(1)$, it follows that $\sin x_n/\sqrt{x_n} = o_p(1)$. All that needs to be shown is that $\sin a_n/\sqrt{a_n} \to 0$ if $a_n \to 0$.

EXAMPLE 3. The following is the simplest of several results which concern Taylor Series Expansions.

COROLLARY 1. *If*

$$x_n = a + o_p(r_n), \tag{1}$$

where $r_n \to 0$, *and*

(2) $f(x)$ *has* s *continuous derivatives at* $x = a$, *then*

$$f(x_n) = f(a) + (x_n - a)f'(a) + \cdots + \frac{(x_n - a)^s f^{(s)}_{(a)}}{s!} + o_p(r_n^s).$$

The following is a considerably more sophisticated example. Here the separation of stochastic and nonstochastic elements is a blessing, for the problem is not completely trivial under the best of circumstances.

EXAMPLE 4. Suppose that $x_1, x_2, \cdots, x_n$ are n independent observations on a chance variable with density

$$f(x \mid \alpha, \beta, \gamma) = \beta \qquad \text{for } 0 \leqq x \leqq \alpha,$$
$$f(x \mid \alpha, \beta, \gamma) = \gamma \qquad \text{for } \alpha < x \leqq 1,$$

where $\alpha\beta + (1 - \alpha)\gamma = 1$, $0 < \alpha < 1$, $\beta > 0$, $\gamma > 0$, and $\beta \neq \gamma$. It is not difficult to show that the maximum likelihood estimate $\hat{\alpha}_n$ of α maximizes

$$\left[\frac{F_n(\alpha)}{\alpha}\right]^{F_n(\alpha)} \left[\frac{1 - F_n(\alpha)}{1 - \alpha}\right]^{1-F_n(\alpha)},$$

where F_n is the sample c.d.f.; i.e., $F_n(x)$ is $1/n$ times the number of observations less than or equal to x. (Note that the function F_n is itself random.) We may write $\hat{\alpha}_n = \Phi(F_n)$. A proof of the consistency of $\hat{\alpha}_n$ (i.e., that $\hat{\alpha}_n \xrightarrow{p} \alpha_0$ if α_0 is the true value of the parameter) is partially complicated by the possibility that

$\hat{\alpha}_n$ may get close to zero or one. We shall merely outline the proof that $\hat{\alpha}_n$ is bounded away from zero in probability, i.e., $1/\hat{\alpha}_n = O_p(1)$.

First, it is known that

$$\sup_{0 \leq x \leq 1} | F_n(x) - F_0(x) | = o_p(1), \tag{1}$$

where $F_0(x)$ is the true c.d.f., and it can be shown that

$$\sup_{0 < x < 1} \left| \frac{F_n(x)}{x} \right| = O_p(1). \tag{2}$$

Secondly, it can be shown that if $\{G_n\}$ is a sequence of nonrandom c.d.f.'s such that

$$\sup_{0 \leq x \leq 1} | G_n(x) - F_0(x) | = o(1), \tag{1'}$$

$$\sup_{0 < x < 1} \left| \frac{G_n(x)}{x} \right| = O(1), \tag{2'}$$

then $1/\Phi(G_n) = O(1)$. It follows that $1/\hat{\alpha}_n = O_p(1)$.

One may observe that the role of x_n in Theorem 1 is played here by the sample c.d.f., F_n.

Another important consideration in the Mann-Wald paper involves a generalization of a well-known result which states that if x_n has a limiting distribution, then for a continuous function g, $g(x_n)$ has the corresponding limiting distribution. Hence, if x_n is asymptotically normally distributed with mean 0 and variance 1, x_n^2 has an asymptotic chi-square distribution with one degree of freedom This result was generalized to allow for the possibility that g has points of discontinuity. Unfortunately, through an oversight, a slightly weaker result than could have been obtained was presented. The stronger version will be stated after we introduce some appropriate notation.

We write $\mathcal{L}(x_n) \to \mathcal{L}(x)$ (read: the distribution law of x_n converges to the distribution law of x) or $\lim_{n\to\infty} \mathcal{L}(x_n) = \mathcal{L}(x)$ if $F_n(a) \to F(a)$ at every point a of continuity of F, where F_n and F are the c.d.f.'s of x_n and x, respectively. Here, $\mathcal{L}(x_n)$ and $\mathcal{L}(x)$ represent the probability measures associated with x_n and x. Let $D(g)$ be the set of discontinuities of the function g.

THEOREM 2. *If*

$$\mathcal{L}(x_n) \to \mathcal{L}(x) \tag{1}$$

and

$$\mathcal{L}(x; D(g)) \equiv P\{x \in D(g)\} = 0, \tag{2}$$

then

$$\mathcal{L}[g(x_n)] \to \mathcal{L}[g(x)].$$

EXAMPLE 1. If $\mathcal{L}(x_n, y_n) \to \mathcal{L}(x, y)$, where x and y are independently and normally distributed with mean 0 and variance 1, then $\mathcal{L}(x_n/y_n) \to \mathcal{L}(x/y)$, which is a Cauchy distribution.

This theorem was extended by Rubin [4] to the case where x_n and x take on values in a topological space X. Here, the notion of convergence in distribution law must be extended. Rubin uses the following definition:[2]

$$\mathfrak{L}_n \to \mathfrak{L} \quad \text{if for every closed set} \quad S, \qquad \mathfrak{L}(S) \geq \limsup_{n \to \infty} \mathfrak{L}_n(S)$$

or, equivalently,

$$\mathfrak{L}_n \to \mathfrak{L} \quad \text{if for every open set} \quad S, \qquad \mathfrak{L}(S) \leq \liminf_{n \to \infty} \mathfrak{L}_n(S).$$

For many spaces, in particular for metric spaces, this definition coincides with the following one used by other authors [5], [6]: $\mathfrak{L}_n \to \mathfrak{L}$ if for every bounded continuous function h,

$$\int h(x)\, d\mathfrak{L}_n(x) \to \int h(x)\, d\mathfrak{L}(x).$$

Both of these are extensions of the definition for Euclidean spaces. With Rubin's definition, it follows rather easily that Theorem 2 applies whenever g is a measurable transformation from one topological space into another.

Rubin [7] has applied this result to find the limiting distribution of quasi-maximum likelihood estimates of the parameters of certain sets of simultaneous linear stochastic difference equations. Donsker [8] derived a related result while engaged in the justification of a heuristic derivation of the asymptotic distribution of Kolmogorov-Smirnov statistic given by Doob [9]. It is interesting to note that in terms of our Theorem 2, Doob's paper dealt mainly with finding the distribution of $g(x)$, after indicating that it seemed reasonable to expect that in some sense $\mathfrak{L}(x_n) \to \mathfrak{L}(x)$. There the role of x_n was played by the sample c.d.f. in the Kolmogorov-Smirnov problem.

The above exposition is far from complete. For example, the following result for Euclidean spaces is rather useful. Note that it can be reworded so as to be extended to metric spaces.

THEOREM 3. *If* $\mathfrak{L}(x_n) \to \mathfrak{L}(x)$, *then* $\mathfrak{L}(x_n + o_p(1)) \to \mathfrak{L}(x)$.

Furthermore, it seems to me that there still remains some work to be done with a view to making the application of Theorems 1 and 2 more cut and dried. Finally, it should be remarked that direct derivations which do not separate the stochastic and asymptotic elements of the problem are sometimes simpler and neater than the techniques suggested by the above results.

3. The Cramér extension of the central limit theorem. In 1938, Cramér [10] obtained an elegant extension of the central limit theorem which, for some reason, seemed to have been overlooked by statisticians. This seems to have been unfortunate, since it appears to be more relevant than the central limit theorem in many statistical applications.

The central limit theorem is loosely described as follows. The average $\bar{X}_n$

[2] The Borel field associated with the distributions is assumed to be that generated by the closed sets.

of n observations on a chance variable X is approximately normally distributed. More precisely,

$$\Pr\left\{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leqq a\right\} \to \int_{-\infty}^{a} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt \qquad \text{as } n \to \infty,$$

if $\bar{X}_n$ is the average of n independent observations on a chance variable with mean μ and variance σ^2. Suppose, now, that a is not fixed but is replaced by a_n, where $a_n \to -\infty$ as $n \to \infty$. Then both sides of the above expression would approach zero. In this sense, the above equation could be considered to be still valid. Even so, it is of importance, as we shall see in Section 6, to determine how fast each side approaches zero and whether the two sides are asymptotically equivalent, i.e., whether the ratio of the two terms approaches one.

In fact, Cramér has essentially shown that as long as a_n does not approach $-\infty$ too rapidly, the two sides are roughly equivalent. However, this result fails to hold when a_n is of the order of magnitude of $\sqrt{n}$. Note that if $a_n = -b\sqrt{n}$, $b > 0$, we are essentially interested in $\Pr\{\bar{X} \leqq c\}$, where $c < \mu$. This case is an especially important one. Here, it is shown that, roughly speaking, $\Pr\{\bar{X} \leqq c\} \approx m^n$, where $m = \inf_t E\{e^{t(X-c)}\}$.

A result of Esseen [11] permits us to eliminate one of the conditions which Cramér had to apply, and which led him to obtain weaker results for the case where the chance variable is discrete. We shall state a version of Cramér's result.

THEOREM 1. *If* $E(e^{tX}) < \infty$ *in some neighborhood of* $t = 0$, *and if* $a_n < -1$, *and* $a_n = o(\sqrt{n})$, *then*

$$\Pr\left\{\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leqq a_n\right\} = \left[\int_{-\infty}^{a_n} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt\right] \cdot \exp\left[\frac{a_n^3}{\sqrt{n}} \lambda\left(\frac{a_n}{\sqrt{n}}\right)\right] \cdot \left[1 + o\left(\frac{a_n}{\sqrt{n}}\right)\right],$$

where $\lambda(t)$ is an analytic function of t whose coefficients depend on the moments of X.

A similar result is obtainable for the case where a_n is positive. Note that

$$(a_n^3/\sqrt{n})\lambda(a_n/\sqrt{n})$$

may become large if a_n is larger in magnitude than $n^{1/6}$. However, this term contributes a relatively unimportant amount compared with the normal approximation term which is asymptotically equivalent to $(\sqrt{2\pi}a_n)^{-1}\exp(-a_n^2/2)$.

THEOREM 2. *If* $E(e^{tX}) < \infty$ *for* t *in some neighborhood of* 0 *and* $c \leqq E(X)$, *then*

$$P\{\bar{X}_n \leqq c\} = \frac{1}{\sqrt{n}} m^n \left[b_0 + \frac{b_1}{n} + \cdots + \frac{b_{k-1}}{n^{k-1}} + O\left(\frac{1}{n^k}\right)\right],$$

where $b_0 > 0$ *and* $m = \inf_t E(e^{t(X-c)})$; the quantities b_i depend on c, and k is an arbitrary positive integer.

Cramér's results were generalized by Feller [12] for the case where the observations do not necessarily have the same distribution.

PART II

4. Estimation. The development of the large-sample theory of estimation was given great impetus with the publication by Fisher [13], [14] of his works on estimation, where he proposed the method of maximum likelihood and suggested, among others, the concepts of consistency, efficiency, and sufficiency. The importance of the notions Fisher developed was soon recognized and the method of maximum likelihood became very popular among statisticians. However, these notions and the properties of the method of maximum likelihood were somewhat more complicated than Fisher or his immediate followers realized. Consequently, many proofs dealing with the properties of these estimates were found to be in error. Considerable light was thrown on these complications when J. L. Hodges, Jr., produced an example of superefficiency. This concept was later treated by Le Cam [15], who also presented an excellent historical survey of the field of maximum likelihood estimation. We shall discuss these notions very briefly, referring the reader to Le Cam's paper for a more detailed discussion.

Let X be a chance variable whose distribution is determined by the value of a parameter θ which is assumed to be in a prescribed set Ω. For the purpose of large-sample theory, Fisher defines an estimate T as a sequence of functions $\{T_n = T_n(X_1, X_2, \cdots, X_n)\}$, where $T_n(X_1, \cdots, X_n)$ represents the "estimated" point of Ω when a sample $X_1, \cdots, X_n$ of n independent observations on X are observed.[3]

DEFINITION 1. T is consistent if $T_n(X_1, \cdots, X_n) \to \theta$ in probability as $n \to \infty$.

Suppose that the distribution of X is characterized by the density $f(x, \theta)$. Then, an estimate T^* is a maximum likelihood estimate of θ if $\prod_{i=1}^{n} f(X_i, \theta)$ assumes its maximum value at $\theta = T_n^*(X_1, X_2, \cdots, X_n)$. (In most applications, the class of distributions may be represented by densities with respect to some σ-finite measure.) It may turn out that the maximum likelihood estimate does not exist. For example, there will be no such estimate for the mean μ of a normal distribution if it is assumed that μ is in the *open* interval $(-1, 1)$ and that the sample mean is greater than one.

When Fisher introduced the notion of asymptotic efficiency, he did this for the case where θ was assumed to be on the real line. Then T was said to be asymptotically efficient if its asymptotic distribution (when properly normalized) was normal with no larger variance than that obtained for any other consistent asymptotically normally distributed statistic. (The variance of the asymptotic distribution will be called the asymptotic variance and is, in general, no larger than the limit of the variance of the normalized estimate.) Apparently, the restriction to asymptotically normally distributed statistics was felt necessary, because Fisher had no way of comparing two dissimilar limiting distributions.

[3] The extension of this notion to the case where the observations need not be independent nor identically distributed is rather evident and we shall not formally treat of that case here.

Fisher and various followers claimed that under suitable mild restrictions the maximum likelihood estimates were consistent and efficient. That the attempts to establish efficiency with the above definition would encounter grave difficulties seems clear when an example of superefficiency is given. Le Cam's example is that of observations from a normal population with unknown mean μ and variance 1. Let T_n represent the maximum likelihood estimate which is the mean of n observations and let T'_n be defined as follows:

$$T'_n = T_n \qquad \text{if} \quad |T_n| \geq \frac{1}{n^{1/4}},$$

$$T'_n = \alpha T_n \qquad \text{if} \quad |T_n| < \frac{1}{n^{1/4}},$$

where α is an arbitrary constant. Then it is clear that[4]

$$\mathcal{L}\{\sqrt{n}(T_n - \mu)\} \to N(0, 1),$$

while

$$\mathcal{L}\{\sqrt{n}(T'_n - \mu)\} \to N(0, 1) \qquad \text{if } \mu \neq 0,$$

but

$$\mathcal{L}\{\sqrt{n}(T'_n - \mu)\} \to N(0, \alpha^2) \qquad \text{if } \mu = 0.$$

Hence, if $0 < \alpha^2 < 1$, T'_n is asymptotically normally distributed with asymptotic variance which is never larger, and sometimes smaller, than that of T_n. Let us call the set of θ, on which a statistic T'_n is more "efficient" than the maximum likelihood estimate T_n, the set of superefficiency. Le Cam has shown under certain conditions that a set of superefficiency must have Lebesgue measure zero. In this sense the maximum likelihood estimate is *efficient*.

In his paper Le Cam makes use of Wald's decision-theory formulation [16] of the estimation problem. (Similar techniques were independently applied by Wolfowitz [17].) His definition of efficiency and superefficiency involves the loss function $L_n(t, \theta)$, which is introduced to represent the loss to the statistician when he observes a sample of size n and estimates t, while θ is the true value of the parameter. Le Cam derives and uses the properties of Bayes' estimates in his attack. I wish to indicate an alternative approach which yields somewhat weaker results but which will be useful to us later. We may assume that L is measured in terms of negative utility [18], so that it makes sense to attempt to select T so as to minimize the "risk" or expected loss $E\{L_n(T_n, \theta)\}$. Then, corresponding to an estimate T, we have a sequence of risk functions

$$R_n(T_n, \theta) = E\{L_n(T_n(X_1, \cdots, X_n), \theta)\}.$$

This formulation permits us to compare estimates which are (1) not necessarily confined to the real numbers and (2) do not necessarily have similar distribu-

[4] $N(0, 1)$ represents the normal distribution with mean 0 and variance 1.

tions. However, a difficulty appears. First of all, it is usually quite difficult to evaluate the loss function that the statistician really faces. On the other hand, in many cases, it is reasonable to assume that $L_n(t, \theta)$ is a minimum at $t = \theta$ and is well behaved near $t = \theta$. Hence, it is often reasonable to assume (in the one-dimensional case) that for t close to θ, $L_n(t, \theta)$ is approximately

$$c_{0n}(\theta) + c_{2n}(\theta)(t - \theta)^2,$$

where $c_{2n}(\theta) > 0$. Intuitively, this would seem to furnish a good excuse for selecting estimates which minimize the second moment about θ. However, some misgivings may arise when we note that $\lim_{n\to\infty} n\{E(T_n - \theta)^2\}$ and the variance of the limiting distribution of $\sqrt{n}(T_n - \theta)$ need not coincide. In extreme cases, it is possible for an estimate to have a very good asymptotic distribution but have infinite variance for each sample size. This estimate would not show up well if we used $E\{(T_n - \theta)^2\}$ as a criterion. In fact, a utility function which satisfies the von Neumann-Morgenstern axioms [18] must be bounded. Hence, $L_n(t, \theta)$ should be taken to be bounded, whereas the above approximation, which may be reasonable for t close to θ, is not. It is difficult to say what is an appropriate criterion without referring to the true $L_n(t, \theta)$. One might propose the asymptotic variance of $T_n - \theta$ (when suitably normalized), but objections could easily be raised against this.

Suppose that one considered estimates T such that

$$T_n - \theta = O_p(1/\sqrt{n}).$$

Let us treat the expectation of the normalized loss function

$$L_n^*(t, \theta) = n\left[\frac{L_n(t, \theta) - c_{0n}(\theta)}{c_{2n}(\theta)}\right],$$

where we assume

$$L_n^*(t, \theta) = n[(t - \theta)^2 + o(t - \theta)^2],$$

and o is assumed to hold uniformly in n as $t \to \theta$. Then,

$$\liminf_{n\to\infty} \frac{E\{L_n^*(T_n, \theta)\}}{nE\left\{\min\left[\left(T_n - \theta\right)^2, \frac{k^2}{n}\right]\right\}} \geq 1;$$

$$\lim_{k\to\infty} \liminf_{n\to\infty} \frac{E\{L_n^*(T_n, \theta)\}}{nE\left\{\min\left[(T_n - \theta)^2, \frac{k^2}{n}\right]\right\}} \geq 1.$$

If $\sqrt{n}(T_n - \theta)$ has a limiting distribution with second moment $\sigma^2(\theta)$, it follows that

$$\lim_{k\to\infty} \lim_{n\to\infty} nE\left\{\left[\min (T_n - \theta)^2, \frac{k^2}{n}\right]\right\} = \sigma^2(\theta)$$

and the asymptotic variance $\sigma^2(\theta)$ may be regarded as a lower bound for the normalized risk function.

On the other hand, if $P\{|T_n - \theta| > k\} = o(1/n)$ for each k, it is possible to show that

$$\liminf_{n\to\infty} \frac{nE\{(T_n - \theta)^2\}}{E\{L_n^*(T_n, \theta)\}} \geqq 1,$$

and then the normalized risk is sandwiched between the real variance (normalized) and the asymptotic variance. A similar discussion is given by Hodges and Lehmann [19].

I believe that without unreasonable modifications the standard derivations of the asymptotic normal distribution of the maximum likelihood estimates can be used to show that for the maximum likelihood estimates $\lim_{n\to\infty} E\{L_n^*(T_n, \theta)\}$ is *equal to the asymptotic variance*. As far as I know, no such proof exists yet in the literature.

The above discussion extends easily to the k-dimensional parameter case where the role of the asymptotic variance is played by an expression of the form $\sum_{i,j} a_{ij}(\theta)\sigma_{ij}(\theta)$ Here $A = \|a_{ij}\|$ is a nonnegative symmetric matrix whose elements correspond to the second-order partial derivatives of the loss function at θ (provided that these derivatives or their ratios converge as $n \to \infty$), and $\|\sigma_{ij}(\theta)\|$ is the asymptotic covariance matrix.

A technique that had been used in previous attempts to establish the efficiency of maximum likelihood estimates was the derivation of a lower bound for the variance of an estimate and the proof that this lower bound was "asymptotically" attained by the maximum likelihood estimates.

Results in this direction were apparently first obtained by Fréchet [20] and Darmois [21] and later given by Cramér [22] and Rao [23] and called the Cramér-Rao inequality. Savage [24] has tentatively suggested alternatively using the name "Information inequality". These results were extended in various directions by Bhattacharya [25], [26], Barankin [27], Wolfowitz [28], Seth [29], Chapman and Robbins [30], Kiefer [31], and Fraser and Guttman [32]. Because these results invoked regularity conditions on the estimates, the possibility of superefficiency was hidden. Let us consider the following form of this result which does not use regularity conditions on the estimates. (This form and a variant of it were communicated to me by Charles Stein and Herman Rubin, respectively.)

First, we consider the *nonasymptotic* case where the parameter space Ω is a subset of the real line containing the origin as an inner point. Let us define Fisher's information by

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f(X, \theta)}{\partial\theta}\right)^2\right],$$

where E_θ represents expectation with respect to the distribution determined by θ. We digress slightly to point out that $I(\theta)$ is additive. That is, if several inde-

pendent observations are combined, the corresponding information is the sum of the individual informations. In particular, when n independent observations are taken on a chance variable X, the information is multiplied by n.

Now let $T(X)$ be an estimate based on the observation X. Under mild conditions on the distribution of X (and not on T) we have

LEMMA 1. *For every* ϵ, $0 < \epsilon < 1$, *and any estimate* T,

$$\sup_{-a<\theta<a} [E_\theta\{(T(X) - \theta)^2\}I(\theta)] \geqq 1 - \epsilon$$

if

$$\frac{1}{2a}\int_{-a}^{a}\frac{d\theta}{I(\theta)} \leqq \frac{a^2\epsilon^2}{12(1-\epsilon)}.$$

Otherwise,

$$\sup_{-a<\theta<a} E_\theta\{(T(X) - \theta)^2\} \geqq \frac{a^2\epsilon^2}{4}.$$

This result can be applied to the large-sample case. To deal with estimates which may behave well asymptotically, but which may have large or even infinite variances, we introduce the truncated estimate T^*,

$$T^* = T \text{ if } |T| \leqq 2a; \qquad T^* = 2a \text{ if } T > 2a; \qquad T^* = -2a \text{ if } T < -2a.$$

Since $\min[(T - \theta)^2, 16a^2] \geqq (T^* - \theta)^2$ for $-a < \theta < a$, we can easily derive the following theorem for the case of n independent observations on X.

THEOREM 1.

$$\lim_{k\to\infty}\liminf_{n\to\infty}\sup_{-k/4\sqrt{n}<\theta<k/4\sqrt{n}} E_\theta\left\{nI(\theta)\min\left[(t-\theta)^2, \frac{k^2}{n}\right]\right\} \geqq 1$$

if $I(\theta)$ *is measurable and bounded away from* 0 *in some neighborhood of* $\theta = 0$. (This statement might be easier to read if it were weakened by replacing, under "sup," the interval $-k/4\sqrt{n} < \theta < k/4\sqrt{n}$ by $-\delta < \theta < \delta$.)

This result clearly allows for the possibility of superefficiency. It is weaker than Le Cam's results, since it does not confine superefficiency to a set of measure zero. On the other hand, this statement fits in very well with our discussion of the normalized risk functions. It states that for an arbitrary estimate *the reciprocal of the information is "essentially" asymptotically a lower bound for the asymptotic variance and hence for the normalized risk function.* This, together with the above-mentioned conjecture that for the maximum likelihood estimate, the normalized risk approaches the asymptotic variance (which coincides with the reciprocal of the information), would give the "essential" efficiency of the maximum likelihood estimate from the normalized risk-function point of view.

Theorem 1 also has the advantage that it can be easily extended to the case where the independent observations are not necessarily from the same population. If the average information per observation is given by

$$\bar{I}_n(\theta) = \frac{1}{n}[I_1(\theta) + \cdots + I_n(\theta)],$$

where $I_j(\theta)$ is the information corresponding to the jth observation or experiment, we can replace $I(\theta)$ in Theorem 1 by $\bar{I}_n(\theta)$, provided $\bar{I}_n(\theta)$ is measurable and

$$\lim_{n\to\infty} \inf_{-k/\sqrt{n}<\theta<k/\sqrt{n}} \bar{I}_n(\theta) > 0$$

for each k. The case where

$$\lim_{n\to\infty} \sup_{-k/\sqrt{n}<\theta<k/\sqrt{n}} \bar{I}_n(\theta) = 0$$

for each k gives no difficulty.

5. Optimal designs for estimating parameters. Suppose that there is available a class of experiments $\{E\}$. A design will consist of a selection of n of these experiments to be performed independently. Suppose that the outcome of each experiment depends only on a real-valued parameter θ which is to be estimated. We shall assume that the true value of θ is approximately known so that it makes sense to consider locally optimal designs. That is to say, that we shall be interested in selecting n experiments so that an estimate of θ, based on the outcomes, will be very good if θ is close to some specified value θ^0.

If n is large, it seems reasonable to select these n experiments, $E_1, E_2, \cdots, E_n$, so as to make the sum of the corresponding informations $\sum_{i=1}^n I(E_i, \theta^0)$ large. If $I(E, \theta^0)$ is maximized by an experiment E_0, it pays to repeat the experiment, E_0, n times. Then, by the Cramér-Rao type of theorem we treated, the asymptotic variance for any design is at least as large as

$$\frac{1}{\bar{I}_n} = \frac{n}{\sum_{i=1}^n I(E_i, \theta^0)} \geqq \frac{1}{I(E_0, \theta^0)},$$

which is the asymptotic variance for the maximum likelihood estimate based on n repetitions of E_0. Furthermore, if the conjecture that for maximum likelihood estimates, the asymptotic variance is equal to the normalized risk is correct, then the normalized risk is asymptotically a minimum for this design.

While the above problem is not very deep, there are certain remarks which are relevant to the extension of this problem to the multidimensional parameter case. First of all, it is quite possible that $I(E, \theta^0)$ does not attain its maximum. A trivial case is the following: Suppose that E_σ corresponds to observing a normal deviate with mean θ and variance σ^2, and suppose that all E_σ are available for $\sigma > 1$. Here, $I(E_\sigma, \theta^0) = 1/\sigma^2$ can be made arbitrarily close to 1 but cannot equal 1. It is apparent that the theoretical difficulty posed by this situation is neither significant nor important.

In general, some experiments are more costly than others, and the formulation involving the selection of a preassigned number of experiments may reasonably be changed to that of selecting an arbitrary number of experiments whose total cost is preassigned. Here, we would attempt to make $\sum I(E_i, \theta^0)$ large,

subject to the restriction $\sum c(E_i) = k$, where $c(E)$ is the cost of the experiment E. Rewriting the above as

$$\sum I(E_i, \theta^0) = \sum \left[\frac{I(E_i, \theta^0)}{c(E_i)}\right] c(E_i),$$

it is evident that we should select that E_0 which maximizes $I(E, \theta^0)/c(E)$, the information per unit cost, and repeat E_0, $k/c(E_0)$ times.

Let us now extend the problem to the following case. Suppose that it is desired to estimate a parameter θ_1, but the distribution of the outcomes of the available experiments depends not only on θ_1, but also on $\theta_2, \cdots, \theta_k$. A special case of this would be that of estimating the slope β of the regression line of Y on x, where $Y = \alpha + \beta x + u$, $\mathcal{L}(u) = N(0, 1)$. Each level x represents an experiment E_x; then let us assume that one has available the set of E_x for which $-1 \leqq x \leqq 1$. It is well known that in this special example the optimal experiment consists of performing E_1 and E_{-1} each half of the time.

To formulate this problem properly, we first note that in the case of k parameters, the information is replaced by the information matrix

$$I(\theta) = \left\| E_\theta \left\{ \frac{\partial \log f(X, \theta)}{\partial \theta_i} \cdot \frac{\partial \log f(X, \theta)}{\partial \theta_j} \right\} \right\|, \qquad i, j = 1, 2, \cdots, k.$$

The information matrix $I(\theta)$ has the additive property; i.e., the information matrix corresponding to the outcome of several independent experiments E_i is equal to the sum of the corresponding information matrices $\sum I(E_i, \theta)$. Another property of interest is the following: Consider the randomized experiment where E_i is performed with probability p_i. Then, the information matrix for the randomized experiment is given by the average $\sum p_i I(E_i, \theta)$.

Let $I_{ij}(\theta)$ represent the (i, j) term of $I(\theta)$ and let $I^{ij}(\theta)$ be the (i, j) term of $I^{-1}(\theta)$. As $1/I(\theta)$ represented the asymptotic variance in the one-dimensional case, so $I^{-1}(\theta)$ represents the asymptotic covariance matrix in the k-dimensional case. In particular, $I^{11}(\theta)$ represents the asymptotic variance of $\sqrt{n}(\hat{\theta}_1 - \theta_1)$.

It now becomes very natural to formulate our problem as being that of selecting n experiments to minimize

$$\left[\sum_{i=1}^{n} I(E_i, \theta^0)\right]^{11}.$$

We may equivalently minimize the (1, 1) element of the inverse of the average information per observation, i.e., we minimize

$$\bar{I}_n^{11}(\theta) = \left[\frac{1}{n} \sum_{i=1}^{n} I(E_i, \theta^0)\right]^{11}.$$

Now the expression on the right-hand side corresponds to the randomized experiment where each E_i is performed with probability $1/n$. By taking n large enough, we can approximate each randomized experiment arbitrarily closely. Hence, we might reformulate our problem as that of selecting that randomized experiment for which $I(E, \theta^0)^{11}$ is minimized.

Each information matrix is nonnegative definite symmetric and may be identified with the point in $k(k + 1)/2$-dimensional space whose coordinates are the elements on and below the main diagonal of the matrix. The class of matrices corresponding to the randomized experiments is the convex set generated by the matrices of the pure experiments. Hence, our problem reduces to that of minimizing a function on a convex set.

I^{11} is a continuous function of I on the set of positive definite symmetric matrices However, I^{11} is not defined for singular matrices. If the distribution of the outcome of an experiment E depended on less than k independent parameters, the information matrix would be singular. Nevertheless, in this case, it can be shown that it would be meaningful to redefine I^{11} by $\lim_{\lambda \to 0+} (I + \lambda A)^{11}$, where A is an arbitrary positive definite symmetric matrix. We then have [33].

THEOREM 1. *If the set R of randomized information matrices $I(\theta^0)$ is closed and bounded, the function $I^{11}(\theta^0)$ attains its maximum on R at a matrix which is a convex combination of $r \leqq k$ of the information matrices corresponding to the nonrandomized experiments.*

This theorem states that there is a locally optimal design for large n which involves at most k of the original experiments. This result considerably reduces the computational problem involved in computing the optimal design. It constitutes a generalization of a similar result by Elfving [34], which applies to linear regression problems with normal deviates. In connection with his result, Elfving indicated an elegant geometrical technique of finding the optimal solution. His technique applies to our more general problem if all the information matrices resemble those of the regression case; i.e., if the typical information matrix for each experiment can be expressed as $\| x_i x_j \|$. In fact, this case occurs quite frequently in applications which are not normal linear regression.

Finally, this result extends to the case where one is interested in estimating s out of the k parameters involved in the experiments. Then the optimal design involves no more than $k + (k - 1) + \cdots + (k - s + 1)$ experiments. This last result is of limited computational applicability if k and s are not small numbers.

6. Testing simple hypotheses. The easiest problem in statistical inference is that of testing a simple hypothesis against a simple alternative. Suppose that the hypothesis H_0 specifies that n independently distributed observations, $X_1, X_2, \cdots, X_n$, have density $f_0(x)$, whereas the alternative H_1 specifies the density $f_1(x)$. It is well known that the class of best tests are the likelihood ratio tests characterized by critical regions which contain all points where the ratio

$$\prod_{i=1}^{n} f_1(X_i) / \prod_{i=1}^{n} f_0(X_i)$$

exceeds some constant c and a subset of those points for which the ratio is equal to c. It is peculiar that in this example, where the small-sample theory is so well understood, the large-sample theory yields resulst of interest.

First, let us note that the above test can be considered to be one that is based on $\bar{Y}_n = 1/n \sum_{i=1}^{n} Y_i$, where $Y_i = \log f_1(X_i)/f_0(X_i)$. But for tests based on

averages of observations, Cramér's results, which were expressed in Section 3, are applicable. These results also apply to tests which are not necessarily likelihood ratio tests. In what follows, we shall assume that Y_i is not necessarily of the above form, but that the test consists of rejecting H_0 if $\bar{Y}_n > a_n$ and that $\mu_0 = E(Y \mid H_0) < E(Y \mid H_1) = \mu_1$.

The probabilities of the two types of error are given by

$$\alpha_n = P\{\bar{Y}_n > a_n \mid H_0\} \qquad \text{and} \qquad \beta_n = P\{\bar{Y}_n \leqq a_n \mid H_0\}.$$

There are several principles which may be invoked for selecting a_n. One of these is that of minimizing $\alpha_n + \lambda\beta_n$ for some $\lambda > 0$. This principle would be especially meaningful if there were an a priori probability ξ, $0 < \xi < 1$, attached to H_0. Then, if l_{ij} represents the loss due to accepting H_i when H_j is correct, the risk would be given by

$$\begin{aligned} R &= \xi l_{00}(1 - \alpha_n) + \xi l_{10}\alpha_n + (1 - \xi)l_{01}\beta_n + (1 - \xi)l_{11}(1 - \beta_n) \\ &= \xi l_{00} + (1 - \xi)l_{11} + \xi(l_{10} - l_{00})\left[\alpha_n + \frac{(1 - \xi)(l_{01} - l_{11})}{\xi(l_{10} - l_{00})}\beta_n\right]. \end{aligned}$$

But for reasonable loss functions, $l_{01} - l_{11}$ and $l_{10} - l_{00}$ are positive. Hence, minimizing R is equivalent to minimizing $\alpha_n + \lambda\beta_n$, where

$$\lambda = \frac{(1 - \xi)(l_{01} - l_{11})}{\xi(l_{10} - l_{00})} > 0.$$

Another situation in which it would be appropriate to use this criterion would be one where it is desired to minimize some function $F(\alpha_n, \beta_n)$, where neither $\partial F(0, 0)/\partial\alpha$ nor $\partial F(0, 0)/\partial\beta$ vanish. Essentially, this boils down to requiring that as $n \to \infty$, α_n and β_n converge to zero at the same rate.

Let

$$m_i(a) = \inf_t E\{e^{t(X-a)} \mid H_i\}, \qquad i = 0, 1,$$

$$\rho(a) = \max\,[m_0(a), m_1(a)], \qquad \rho = \inf_{\mu_0 \leqq a \leqq \mu_1} \rho(a).$$

A consequence of Cramér's result (see [35]) is

THEOREM 1.

$$\lim_{n\to\infty} [\inf_{a_n} (\beta_n + \lambda\alpha_n)]^{1/n} = \rho \ (\textit{independent of } \lambda).$$

This theorem permits us to compare the relative efficiency of two tests. For the above test, $\beta_n + \lambda\alpha_n$ behaves roughly like ρ^n. Suppose that a similar test is based on the average of another statistic Z. If ρ^* is the corresponding value of ρ for this new test, then

$$e = \frac{\log \rho^*}{\log \rho}$$

is a reasonable measure of the relative efficiency of the test based on Z to the test based on Y. The reason for this is that if n_1 and n_2 are large sample sizes for which the $\alpha_{n_i} + \lambda\beta_{n_i}$ of the two tests are approximately equal, then n_1/n_2 is close to e. In other words, the first test requires en_2 observations to do as well as the second. This measure of efficiency permits us not only to compare various tests based on a given experiment, but also permits us to compare tests based on different experiments.

In particular, let us consider the likelihood ratio test for a given experiment. We designate the corresponding ρ by ρ_{LR}, which can be shown to be given by

$$\rho_{LR} = \inf_{0<t<1} \int [f_1(x)]^t [f_0(x)]^{1-t}\, d\nu(x)$$

if $f_1(x)$ and $f_0(x)$ are the densities of X, with respect to the measure ν, under H_1 and H_0, respectively. Because of the character of the above-mentioned measure of relative efficiency, it is natural to define the information in the experiment by

$$I = -\log \rho_{LR}.$$

Fisher's measure of information also had the property that if two experiments yield informations $I_1(\theta)$ and $I_2(\theta)$, where $I_1(\theta) = 2I_2(\theta)$, then one needs approximately $2n$ observations on the second experiment to get results comparable to those obtained with n observations on the first experiment for n large. It is interesting to note that while Fisher's measure of information is additive, the above is not. In fact, it has the following properties:

(1) The information derived from n independent observations on a chance variable is n times the information from one observation.

(2) The information derived from observations on several independent chance variables is less than or equal to the sum of the corresponding informations.

It occasionally happens in practice that it is important to obtain β very small, whereas a relatively large value of α, like .05 or .10, is not disastrous. In such cases, it makes sense to consider in our large-sample approach the problem where one minimizes β subject to fixed α. Let β_n^* be the value of β_n which corresponds to a fixed value of α, say α_0, $0 < \alpha_0 < 1$. We have, as another consequence of Cramér's result,

THEOREM 2.

$$\lim_{n\to\infty} \beta_n^{*1/n} = \rho^* = m_1(\mu_0) \qquad (\textit{independent of } \alpha_0),$$

where $\mu_0 = E(Y \mid H_0)$.

In particular, for the likelihood ratio test, it is easy to show that we obtain ρ_{LR}^*, which is given by

$$\rho_{LR}^* = m_1(\mu_0) = e^{\mu_0} = \exp\left[\int f_0(x) \log \frac{f_1(x)}{f_0(x)}\, d\nu(x)\right].$$

This result was first obtained by Charles Stein [36]. Here again, it makes sense to define a corresponding measure of information by

$$I^* = -\log \rho^*_{LR} = -\int \log\left[\frac{f_1(x)}{f_0(x)}\right] f_0(x)\, d\nu(x).$$

It is of interest to note that I^* represents one of the Kullback-Leibler information numbers [37]; also

$$I^{**} = \int f_1(x) \log \frac{f_1(x)}{f_0(x)}\, d\nu(x)$$

would arise naturally if β_n were kept fixed and $\alpha_n \to 0$. The Kullback-Leibler numbers do have the additive property. Incidentally, the above characterization of the Kullback-Leibler numbers implies that they do exceed $I = -\log \rho_{LR}$.

Until now, we have not discussed sequential analysis from a large-sample point of view. At a first naive glance, it may seem as though the very nature of sequential analysis is such as to rule out large-sample theory. That this is not so becomes clear when one considers that reducing the cost of sampling should increase the expected sample size. In fact, let us suppose that the cost per observation is c. Consider the Bayes procedure corresponding to a fixed a priori probability ξ that H_0 is correct. The risk function is given by

$$R_0 = l_{00} + \alpha(l_{10} - l_{00}) + cE(n \mid H_0),$$
$$R_1 = l_{11} + \beta(l_{01} - l_{11}) + cE(n \mid H_1).$$

The Bayes risk, $\xi R_0 + (1 - \xi)R_1$, is minimized by Wald's sequential probability ratio test [38]. As $c \to 0$, $E(n \mid H_0)$ and $E(n \mid H_1) \to \infty$, but

$$\xi(R_0 - l_{00}) + (1 - \xi)(R_1 - l_{11}) \to 0.$$

An elementary application of Wald's inequalities concerning the operating characteristic function gives

THEOREM 3.

$$\lim_{c\to 0} \frac{R_0 - l_{00}}{(c \log 1/c)} = \frac{1}{I^*}; \qquad \lim_{c\to 0} \frac{R_1 - l_{11}}{(c \log 1/c)} = \frac{1}{I^{**}}.$$

Note that these limits do not depend on $l_{10} - l_{00}$ nor on $l_{01} - l_{11}$. This is due to the fact that as $c \to 0$, the main part of the risk is the cost of sampling.

It is rather striking that the notions of information, which are natural for the sequential and nonsequential cases, are not identical. Upon some consideration, however, it is not surprising. In the sequential case, after many observations are taken, one is almost sure which hypothesis is correct. Then if H_0 seems correct, the remaining observations may be selected from an experiment for which the corresponding Kullback-Leibler information I^* is large. In the nonsequential case, the experiment to be performed must be decided on before any data are taken. It is natural that the corresponding information should differ from I^* and I^{**}.

It is of interest to note that as the hypotheses H_0 and H_1 get closer to one another, the three measures of information behave in the following fashion:

$$I \approx \frac{I^*}{4} \approx \frac{I^{**}}{4}.$$

7. Composite hypotheses. A classical result in the large-sample theory applied to tests of composite hypotheses is that of Wilks [39]. It states that[5]

$$\mathcal{L}(-2 \log \lambda_n) \to \mathcal{L}(\chi^2_{k-r})$$

if λ_n is the likelihood ratio based on n independent observations for the test that a parameter θ lies on a specified r-dimensional hyperplane of k-dimensional space and the hypothesis is true. It is striking that this result does not involve the distribution of the data except in that mild regularity conditions on the distribution are required.

Many tests of composite hypotheses are not of this simple form. For example, it may be desired to test whether θ lies in the first quadrant of the plane, or it may be desired to test whether θ lies above a hyperplane or even whether θ lies inside a sphere.

For these problems, first suggested to me by Leonid Hurwicz, a natural generalization of Wilks' result is easily obtained.

Let $f(x, \theta)$ represent the density of the data. Suppose that θ lies in k-dimensional space and let ω and τ be two disjoint subsets of this space. We are interested in testing $H_0: \theta \varepsilon \omega$ against the alternative $H_1: \theta \varepsilon \tau$. Let

$$P_\omega(X_1, X_2, \cdots, X_n) = \sup_{\theta \varepsilon \omega} \sum_{i=1}^{n} f(X_i, \theta).$$

The standard definition of the likelihood ratio is given by

$$\lambda_n = \frac{P_\omega(X_1, \cdots, X_n)}{P_{\omega \cup \tau}(X_1, X_2, \cdots, X_n)}.$$

It is somewhat more convenient to treat a more symmetric form

$$\lambda_n^* = \frac{P_\omega(X_1, X_2, \cdots, X_n)}{P_\tau(X_1, X_2, \cdots, X_n)}.$$

These are related by

$$\lambda_n = \lambda_n^* \text{ if } \lambda_n^* \leqq 1; \qquad \lambda_n = 1 \text{ if } \lambda_n^* > 1.$$

We call a set C positively homogeneous if $X \varepsilon C$ implies $aX \varepsilon C$ for all $a > 0$. We say that ω is approximated by a positively homogeneous set C_ω if

$$\inf_{x \varepsilon C_\omega} |x - y| = o(|y|) \qquad \text{for} \quad y \varepsilon \omega,$$

and

$$\inf_{y \varepsilon \omega} |x - y| = 0(|x|) \qquad \text{for} \quad x \varepsilon C_\omega.$$

Let $I(\theta)$ represent Fisher's information matrix.

[5] χ^2_{k-r} represents chi-square with $k - r$ degrees of freedom.

Under certain regularity conditions on $f(x, \theta)$, we have the following result [40]:

THEOREM 1. *If ω and τ are approximated by two disjoint positively homogeneous sets C_ω and C_τ and the true value of θ is at the origin, then the distribution of $-2 \log \lambda_n^*$ is the same as it would be for the case where $\mathcal{L}(X_i) = N(\theta, I(0)^{-1})$ and ω and τ are replaced by C_ω and C_τ.*

The advantage of this result lies in the fact that the case of normally distributed data is relatively simple to treat.

It is now easy to show that if ω is a smooth r-dimensional surface and τ is the rest of the k dimensional space and $\theta \varepsilon \omega$, then

$$\lim \mathcal{L}(-2 \log \lambda_n) = \lim \mathcal{L}(-2 \log \lambda_n^*) = \mathcal{L}(\chi_{k-r}^2).$$

It is also easy to show that if ω is the set on one side of a smooth $(k - 1)$-dimensional surface, τ is the rest of k-dimensional space, and θ is on the boundary,

$$\lim_{n\to\infty} \mathcal{L}(-2 \log \lambda_n^*) = \mathcal{L}(u\chi_1^2); \qquad \lim_{n\to\infty} \mathcal{L}(-2 \log \lambda_n) = \mathcal{L}(v\chi_1^2),$$

then where u is independent of χ_1^2 and takes on the values 1 and -1 with probability $\frac{1}{2}$ and $v = \frac{1}{2}(u + 1)$.

In particular, this case applies to testing whether θ lies inside or outside a sphere, and to testing whether θ lies above or below a hyperplane.

In the problem where one is interested in whether θ is in the first quadrant or not, the following is the situation. If θ is on the positive part of either axis,

$$\lim_{n\to\infty} \mathcal{L}(-2 \log \lambda_n^*) = \mathcal{L}(u\chi_1^2).$$

If θ is at the origin, the limiting distribution depends on $I(0)$ and is not difficult to evaluate mumerically.

8. Summarizing remarks. The topic of this paper is so broad and current research in it is so vigorous that it is impossible for me to do more than mention a few of those notions in it that have been of special interest to me. I have tried to give some feeling for those aspects which attract me to the subject and, in so doing, I have neglected a considerable amount of important work done by many people including among others Neyman and Wald.

REFERENCES

[1] W. G. COCHRAN, "The χ^2 distribution for the binomial and Poisson series, with small expectations," *Ann. Eugenics*, Vol. 7 (1936), pp. 207–217.

[2] H. CHERNOFF AND H. RUBIN, "Asymptotic properties of limited-information estimates under generalized conditions," *Studies in Econometric Method*, William C. Hood and L. C. Koopmans, (eds.) Cowles Commission Monograph 14, John Wiley and Sons, New York, 1953, pp. 200–212.

[3] H. B. MANN AND A. WALD, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[4] H. RUBIN, "Convergence of probability measures on completely regular spaces," unpublished.

[5] B. V. GNEDENKO AND A. N. KOLMOGOROFF, *Limit distributions for sums of independent random variables*, trans. by K. L. Chung, Addison-Wesley, Cambridge, Mass., 1954, 264 pp.

[6] M. FRECHET, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Ann. Inst. H. Poincaré*, Vol. 10 (1948), pp. 215-230.

[7] H. RUBIN, "Systems of linear stochastic equations," unpublished Ph.D. dissertation, University of Chicago, 1948, 50 pp.

[8] M. D. DONSKER, "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277-281.

[9] J. L. DOOB, "Heuristic approach to the Kolmogoroff-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20, (1949), pp. 393-403.

[10] H. CRAMÉR, "Sus un nouveau théorème—limite de la théorie des probabilitás," *Actualités Sci. Ind.*, No. 736, Paris (1938).

[11] C. G. ESSEEN, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1-125.

[12] W. FELLER, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361-372.

[13] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London*, Series A, Vol. 222 (1922), pp. 390-368.

[14] R. A. FISHER, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22, Part 5 (1925), pp. 700-725.

[15] L. LE CAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *Univ. California Publ. Stat.*, Vol. 1 (1953), pp. 277-330.

[16] A. WALD, *Statistical decision functions*, John Wiley and Sons, New York, 1950, 179 pp.

[17] J. Wolfowitz, "The method of maximum likelihood and the Wald theory of decision functions," *Proc. Roy. Dutch Acad. Sci.*, Vol. 56 (1953), pp. 114-119.

[18] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, 2d ed., Princeton University Press, Princeton, N. J,. 1947, 641 pp.

[19] J. L. HODGES AND E. L. LEHMANN, "The Robbins mono-process in the bounded case," to be published in the *Third Berkeley Symposium Proceedings*, University of California Press.

[20] M. FRÉCHET, "Sur l'extension de certains évaluations statistiques au cas de petits échantillons", *Revue de L'Institut International de Statistique*, 11 (1943), pp. 183-205.

[21] G. DARMOIS, "Sur les limites de la dispersion de certains estimations", *Revue de l'Institut International de Statistique*, 13 (1945), pp. 9-15.

[22] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946, 575 pp.

[23] C. R. RAO, "Information and accuracy obtainable in one estimation of a statistical parameter," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81-91.

[24] L. J. SAVAGE, The Foundations of Statistics, John Wiley and Sons, Inc., New York (1954), 294 pp.

[25] A. BHATTACHARYA, "On some analogues of the amount of information and their uses in statistical estimation," *Sankhyā*, Vol. 8 (1946), pp. 1-14.

[26] A. BHATTACHARYA, "On some analogues of the amount of information and their uses in statistical estimation," *Sankhyā*, Vol. 8 (1947), pp. 201-218.

[27] E. W. BARNAKIN, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477-501.

[28] J. WOLFOWITZ, "Efficiency of sequential estimates," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215-230.

[29] G. R. SETH, "On the variance of estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 1-27.

[30] D. G. CHAPMAN AND H. ROBBINS, "Minimum variance estimation without regularity assumptions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 581-586.

[31] J. Kiefer, "On minimum variance estimators," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 627–628.
[32] D. A. S. Fraser and I. Guttman, "Bhattacharya bounds without regularity conditions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 629–631.
[33] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.
[34] G. Elfving, "Optimum allocation in linear regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–262.
[35] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.
[36] C. Stein, "Information and comparison of experiments, unpublished.
[37] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 79–86.
[38] A. Wald, *Sequential Analysis*, John Wiley and Sons, New York, 1947, pp. 60–62.
[39] S. S. Wilks, "The large sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 60–62.
[40] H. Chernoff, "On the distribution of the likelihood ratio," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 573–578.

A "MIXED MODEL" FOR THE ANALYSIS OF VARIANCE[1]

By Henry Scheffé

University of California, Berkeley

1. Summary. A "mixed model" is proposed in which the problem of the appropriate assumptions to make about the joint distribution of the random main effects and interactions is solved by letting this joint distribution follow from more basic and "natural" assumptions about the cell means. The expectations of the mean squares ordinarily calculated turn out, with suitable definition of the variance components, to have the same values as those usually found in more restrictive models, and some of the customary tests and confidence intervals are justified, but some aspects appear to be novel. For example, the over-all test found for the fixed main effects and the associated multiple-comparison method require Hotelling's T^2.

2. Introduction. We consider K replications of a two-way layout with I rows and J columns ($I > 1$, $J > 1$, $K \geqq 1$), the rows corresponding to levels of a "Model I" [4] factor A, whose effects we wish to regard as fixed effects, and the columns corresponding to the levels of a "Model II" factor B, whose effects we wish to regard as random effects. We let y_{ijk} denote the kth measurement in the i, j cell (the intersection of the ith row and jth column). Throughout this paper, i and i', as subscripts or indices of summation, will range over the integers from 1 to I; j, j', and j'' will range from 1 to J, etc., unless otherwise indicated.

As an illustration, we may imagine an experiment involving I different makes of machines and J workers regarded as a sample from a large population of workers. Each worker is put on each machine for K days during the experiment and y_{ijk} is a measurement of the output of the jth worker the kth day he is on the ith machine. It is customary in the analysis of variance to write

$$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}, \tag{1}$$

where the general mean μ and the row effects $\{\alpha_i\}$ are constants, about which we may assume without loss of generality that $\sum_i \alpha_i = 0$, and where the column effects $\{b_j\}$, interactions $\{c_{ij}\}$, and "errors" $\{e_{ijk}\}$ are random variables about whose joint distribution certain assumptions are made. The usual assumption that the set $\{e_{ijk}\}$ is statistically independent of the set $\{b_j, c_{ij}\}$ seems acceptable to the writer in many applications, but the further assumptions usually made on the $\{b_j\}$ and $\{c_{ij}\}$ seem to him unsatisfactory,[2] as being *ad hoc*, or too restrictive, or not sufficiently complete. For example, the usual assumption that the $\{c_{ij}\}$ are statistically independent of the $\{b_j\}$ is *ad hoc*, the frequent assumption that

Received February 11, 1955.

[1] This paper was prepared with the partial support of the Office of Naval Research. It may be reproduced in whole or in part for any purpose of the United States government.

[2] An example is [12].

all the $\{c_{ij}\}$ are independent is too restrictive, and the additional assumptions stated by those who take $\sum_i c_{ij} = 0$ (all j) are sometimes insufficient even to determine the expected values of the mean squares usually employed.

3. The model. We propose to avoid the unpleasant assumptions as follows: We will assume that

$$y_{ijk} = m_{ij} + e_{ijk}, \tag{2}$$

where the set of errors $\{e_{ijk}\}$ is statistically independent of the set $\{m_{ij}\}$ of "true" cell means.

About the set of errors $\{e_{ijk}\}$, we assume that they are independently and identically distributed with zero means and variance σ_e^2. (This assumption can obviously be lightened without affecting the validity of the expected mean squares and unbiased estimates derived in Section 5, which depends only on the first and second moments of the set $\{m_{ij}, e_{ijk}\}$, and for which it is sufficient that the set $\{e_{ijk}\}$ have zero means, zero correlations, and a common variance σ_e^2.)

The writer hopes that the assumptions about the joint distribution of the set $\{m_{ij}\}$, to be stated below, will be found acceptable. The main effects $\{b_j\}$ and interactions $\{c_{ij}\}$ will then be defined in terms of the $\{m_{ij}\}$ in a natural way, and the joint distribution of the set $\{b_j, c_{ij}\}$ will thus be determined. Some parts of this program and its implications have also been developed earlier by others, as we shall be able to indicate more conveniently at the end of this paper.

Our basic assumption on the rectangular array of the $\{m_{ij}\}$ is that the J columns are distributed independently like a vector random variable m with I components, $m_1, \cdots, m_I$. Thus, in the above illustration of machines and workers, for each worker in the population there is a vector whose I components are his "true means" on the I machines, and the random vector m has the distribution generated by this population of workers, idealized as an infinite population. The J columns of the array $\{m_{ij}\}$ are the J vectors belonging to the J workers in the experiment, assumed to be a random sample from the population of workers.

About the components $\{m_i\}$ of the random vector m, we shall always assume that the I variances are finite. Sometimes we shall also make the normality assumption

($\mathfrak{N}$): The $\{m_i\}$ have a joint normal distribution, and the $\{e_{ijk}\}$ are also jointly normal.

We shall also have occasion to refer to a symmetry assumption

($\mathcal{S}$): The $\{m_i\}$ have equal variances and equal covariances.

We will refer to the assumption ($\mathcal{S}$) as a limiting case in which certain relations become simpler or clearer, but we do *not* recommend it in applications—where there is usually no real symmetry corresponding to this assumption. Thus, in our illustration, two machines might be very similar (perhaps of the same make and model), but very different from the other machines. Further objections to assuming ($\mathcal{S}$) will arise when we consider below the finite analogue of the infinite population of vectors associated with the random vector m.

4. Definition of effects and variance components. The "true" mean for the ith level of the factor A (the reader may find it easier to substitute " 'true' mean for the ith machine") is defined to be

$$\mu_i = E(m_i), \tag{3}$$

and the "true" general mean, to be

$$\mu = \mu_. , \tag{4}$$

where the dot notation here and elsewhere signifies that the arithmetic average has been taken over the subscript which the dot replaces, that is, $\mu_. = \sum_i \mu_i / I$. The main effect of the ith level of A, or the ith row effect, is defined to be

$$\alpha_i = \mu_i - \mu, \tag{5}$$

so that $\sum_i \alpha_i = 0$. The "true" meae for the jth level of factor B (for the jth worker in the experiment) is defined to be $m_{.j}$, and the main effect of the jth level of B, or the jth column effect, to be

$$b_j = m_{.j} - \mu. \tag{6}$$

Finally, the interaction effect, c_{ij}, of the jth level of B with the ith level of A is defined to satisfy the equation

$$m_{ij} = \mu + \alpha_i + b_j + c_{ij}, \tag{7}$$

or

$$c_{ij} = m_{ij} - m_{.j} - \alpha_i, \tag{8}$$

so that $\sum_i c_{ij} = 0$ (all j).

We see that the J vectors with $I + 1$ components $b_j, c_{1j}, \cdots, c_{Ij}$ are independently distributed like the random vector with components $b, c_1, \cdots, c_I$ defined in terms of the basic vector m as follows:

$$b = m. - \mu, \tag{9}$$

$$c_i = m_i - m. - \alpha_i. \tag{10}$$

We note that b and the $\{c_i\}$ have zero means, and that their variances and covariances depend on the elements of the covariance matrix $(\sigma_{ii'})$ of the vector m in the following way:

$$\text{var}(b) = \sigma_{..}, \tag{11}$$

$$\text{cov}(c_i, c_{i'}) = \sigma_{ii'} - \sigma_{i.} - \sigma_{i'.} + \sigma_{..}, \tag{12}$$

$$\text{cov}(b, c_i) = \sigma_{i.} - \sigma_{..}. \tag{13}$$

The main effects $\{b_j\}$ and interactions $\{c_{ij}\}$ in (1) thus have zero means, and the variances and covariances within a set $\{b_j, c_{1j}, \cdots, c_{Ij}\}$ are given by (11), (12), (13), while the covariance of any member of this set with any member of the set $\{b_{j'}, c_{1j'}, \cdots, c_{Ij'}\}$ is zero for $j \neq j'$.

We shall be led to the appropriate definitions of the "variance components" σ_A^2, σ_B^2, σ_{AB}^2, by way of the analogy with the "finite model", where the vector m can take on only one of a finite number, Q, of values, the qth having components, say, $\mu_{1q}, \cdots, \mu_{Iq}$. For the corresponding $I \times Q$ rectangular array, the usual definitions, chosen to give the simplest formulas for the expectations of the mean squares customarily computed, are

$$\sigma_A^2 = (I - 1)^{-1} \sum_i (\mu_{i\cdot} - \mu_{\cdot\cdot})^2, \tag{14}$$

$$\sigma_B^2 = (Q - 1)^{-1} \sum_q (\mu_{\cdot q} - \mu_{\cdot\cdot})^2, \tag{15}$$

$$\sigma_{AB}^2 = (I - 1)^{-1} (Q - 1)^{-1} \sum_i \sum_q (\mu_{iq} - \mu_{i\cdot} - \mu_{\cdot q} + \mu_{\cdot\cdot})^2. \tag{16}$$

If we regard our previous infinite model as a limiting case of the finite model as $Q \to \infty$, we see that the analogues of the above formulas are to be found by replacing μ_{iq}, $\mu_{i\cdot}$, $\mu_{\cdot q}$, $\mu_{\cdot\cdot}$, $(Q - 1)^{-1}\sum_q$ by m_i, μ_i, $m.$, μ, E, respectively, and we are led to the following definitions, which we shall adopt for the infinite model:

$$\sigma_A^2 = (I - 1)^{-1} \sum_i \alpha_i^2, \tag{17}$$

$$\sigma_B^2 = \text{var}\,(b), \tag{18}$$

$$\sigma_{AB}^2 = (I - 1)^{-1} \sum_i \text{var}\,(c_i). \tag{19}$$

The variance components σ_B^2 and σ_{AB}^2 may be expressed in terms of the elements of the covariance matrix $(\sigma_{ii'})$ of the vector m, from (11) and (12),

$$\sigma_B^2 = \sigma_{\cdot\cdot}, \tag{20}$$

$$\sigma_{AB}^2 = (I - 1)^{-1} \sum_i (\sigma_{ii} - \sigma_{\cdot\cdot}). \tag{21}$$

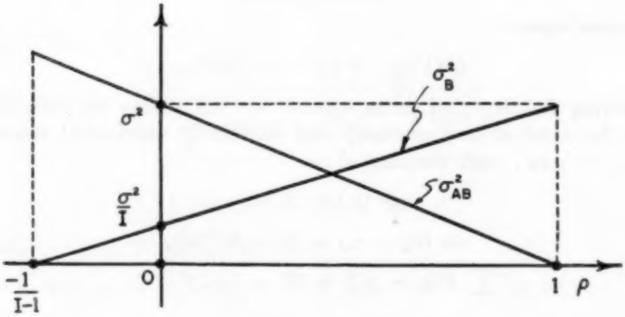
We note that $\sigma_B^2 = 0$ if and only if $b = 0$ (we omit the phrase "with probability one" here and elsewhere where it obviously applies), that is, if and only if the basic vector m has a degenerate distribution satisfying $\sum_i m_i = \text{constant} = I\mu$. Also, $\sigma_{AB}^2 = 0$ if and only if var $(c_i) = 0$ for all i, or $c_i = 0$ for all i, or $m_i = m. + \alpha_i$; that is, except for additive constants $\{\alpha_i\}$, the random variables m_i are *identical* (not just identically distributed). Some further insight into our definitions of the random main and interaction effects and their variance components may be obtained by considering the symmetric case $(\mathcal{S})$ where $\sigma_{ii'} = \rho\sigma^2$ if $i \neq i'$, $\sigma_{ii} = \sigma^2$. Then, from (20) and (21),

$$\sigma_B^2 = \sigma^2 [1 + \rho(I - 1)]/I, \tag{22}$$

$$\sigma_{AB}^2 = \sigma^2 (1 - \rho), \tag{23}$$

where $-(I - 1)^{-1} \leqq \rho \leqq 1$. These relations are shown graphically in Fig. 1.

The previously mentioned objection to assuming $(\mathcal{S})$ in the infinite model is



Variance components σ^2_B and σ^2_{AB} in symmetric case (S)

FIG. 1

that its analogue in the finite model is the fulfillment of the following $\frac{1}{2}I(I+1)-2$ conditions: If $g_{ii'}$ denotes

(24) $$\sum_q (\mu_{iq} - \mu_{i\cdot})(\mu_{i'q} - \mu_{i'\cdot}),$$

then all g_{ii} are equal, and all $g_{ii'}$ with $i \neq i'$ are equal. There would seem to be nothing in most applications to justify this.

5. Expected mean squares and point estimates. We shall consider the customary sums of squares—namely, those for rows, columns, interaction, and error, which we shall denote by $(SS)_A$, $(SS)_B$, $(SS)_{AB}$, $(SS)_e$, respectively—and the corresponding mean squares,

(25) $$(MS)_A = (I-1)^{-1}(SS)_A,$$

(26) $$(MS)_B = (J-1)^{-1}(SS)_B,$$

(27) $$(MS)_{AB} = (I-1)^{-1}(J-1)^{-1}(SS)_{AB},$$

(28) $$(MS)_e = I^{-1}J^{-1}(K-1)^{-1}(SS)_e,$$

where

(29) $$(SS)_A = JK\sum_i (y_{i\cdot\cdot} - y_{\cdot\cdot\cdot})^2,$$

(30) $$(SS)_B = IK\sum_j (y_{\cdot j\cdot} - y_{\cdot\cdot\cdot})^2,$$

(31) $$(SS)_{AB} = K\sum_i \sum_j (y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdot\cdot\cdot})^2,$$

(32) $$(SS)_e = \sum_i \sum_j \sum_k (y_{ijk} - y_{ij\cdot})^2.$$

In addition, we shall need the contribution to $(SS)_{AB}$ from the ith row,

(33) $$(SS)_{AB,i} = K\sum_j (y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdot\cdot\cdot})^2,$$

and its mean square

$$(MS)_{AB,i} = (J - 1)^{-1}(SS)_{AB,i}. \tag{34}$$

In deriving the expected mean squares we will utilize the following three formulas for a set of independently and identically distributed random variables, $x_1, \cdots, x_N$, with variance σ_x^2:

$$\text{var}(x.) = N^{-1}\sigma_x^2, \tag{35}$$

$$\text{var}(x_n - x.) = (1 - N^{-1})\sigma_x^2, \tag{36}$$

$$\sum_n E(x_n - x.)^2 = (N - 1)\sigma_x^2. \tag{37}$$

It is convenient to define now

$$\hat{\alpha}_i = y_{i\cdots} - y_{\cdots}. \tag{38}$$

We have from (1),

$$\hat{\alpha}_i = \alpha_i + c_{i\cdot} + e_{i\cdot\cdot} - e_{\cdots}, \tag{39}$$

since $c._j = 0$ and hence $c.. = 0$. It follows that

$$E(\hat{\alpha}_i) = \alpha_i \tag{40}$$

and

$$\begin{aligned} \text{var}(\hat{\alpha}_i) &= \text{var}(c_{i\cdot}) + \text{var}(e_{i\cdot\cdot} - e_{\cdots}) \\ &= J^{-1}\,\text{var}(c_i) + (1 - I^{-1})\,\text{var}(e_{i\cdot\cdot}), \end{aligned} \tag{41}$$

from (35) and (36). Again from (35),

$$\text{var}(\hat{\alpha}_i) = J^{-1}\,[\text{var}(c_i) + K^{-1}(1 - I^{-1})\sigma_e^2]. \tag{42}$$

Writing

$$(SS)_A = JK\sum_i \hat{\alpha}_i^2, \tag{43}$$

we may substitute (42) in

$$E(SS)_A = JK\sum_i [\text{var}(\hat{\alpha}_i) + \alpha_i^2] \tag{44}$$

to get

$$E(SS)_A = K\sum_i \text{var}(c_i) + (I - 1)\,\sigma_e^2 + JK\sum_i \alpha_i^2. \tag{45}$$

Using the definitions (17) and (19), we then find that

$$E(MS)_A = JK\sigma_A^2 + K\sigma_{AB}^2 + \sigma_e^2. \tag{46}$$

After substituting (1) into (30) we have

$$(SS)_B = IK\sum_j (b_j - b. + e._{j\cdot} - e_{\cdots})^2, \tag{47}$$

and so from (37),

$$(48)\qquad E(SS)_B = IK(J-1)\operatorname{var}(b_j + e_{.j.}) = IK(J-1)(\sigma_B^2 + I^{-1}K^{-1}\sigma_e^2);$$

hence

$$(49)\qquad E(MS)_B = IK\sigma_B^2 + \sigma_e^2.$$

Substitution of (1) in (33) gives

$$(50)\qquad (SS)_{AB,i} = K\sum_j (c_{ij} - c_{i.} + e_{ij.} - e_{i..} - e_{.j.} + e_{...})^2;$$

whence

$$(51)\qquad E(SS)_{AB,i} = K\sum_j E(c_{ij} - c_{i.})^2 + K\sum_j E(e_{ij.} - e_{i..} - e_{.j.} + e_{...})^2.$$

Call the last term a_i. It is clear that the value of a_i does not depend on i, and it is known from Model I theory that $\sum_i a_i = (I-1)(J-1)\sigma_e^2$. Thus, $a_i = (1 - I^{-1})(J-1)\sigma_e^2$. By (37), the first term on the right of (51) may be written $K(J-1)\operatorname{var}(c_i)$. Hence,

$$(52)\qquad E(SS)_{AB,i} = (J-1)[K\operatorname{var}(c_i) + (1 - I^{-1})\sigma_e^2],$$

$$(53)\qquad E(MS)_{AB,i} = K\operatorname{var}(c_i) + (1 - I^{-1})\sigma_e^2.$$

Summing (52) over i and dividing by $(I-1)(J-1)$, we get

$$(54)\qquad E(MS)_{AB} = K\sigma_{AB}^2 + \sigma_e^2.$$

Finally, if we rewrite (32) as

$$(55)\qquad (SS)_e = \sum_i\sum_j\sum_k (e_{ijk} - e_{ij.})^2,$$

we see that for $K > 1$

$$(56)\qquad E(MS)_e = \sigma_e^2.$$

We shall use the noun "estimate" always to mean "unbiased estimate." The above formulas for the expected mean squares lead to the following estimates of σ_B^2, σ_{AB}^2, σ_e^2, respectively, if $K > 1$:

$$(57)\qquad \hat{\sigma}_B^2 = (IK)^{-1}[(MS)_B - (MS)_e],$$

$$(58)\qquad \hat{\sigma}_{AB}^2 = K^{-1}[(MS)_{AB} - (MS)_e],$$

$$(59)\qquad \hat{\sigma}_e^2 = (MS)_e.$$

An estimate of α_i is the $\hat{\alpha}_i$ defined by (38); an estimate of its variance (42) is $J^{-1}K^{-1}(MS)_{AB,i}$. An estimate of $\mu_i = \mu + \alpha_i$ is $y_{i..}$; an estimate of its variance is $J^{-1}\hat{\tau}_{ii}$, where $\hat{\tau}_{ii}$ is defined by (62) below. An estimate of $\alpha_i - \alpha_{i'}$ is $y_{i..} - y_{i'..}$; an estimate of its variance is

(60) $$S_{ii'} = J^{-1}(J-1)^{-1}\sum_j (y_{ij\cdot} - y_{i'j\cdot} - y_{i\cdot\cdot} + y_{i'\cdot\cdot})^2.$$

In order to estimate the covariance matrix $(\sigma_{ii'})$ of the basic vector m, we note that the J columns of cell means $\{y_{ij\cdot}\}$ are distributed independently like a random vector $u = m + v$, where v is statistically independent of m and has the distribution of the vector with components $e_{1j\cdot}, \cdots, e_{Ij\cdot}$ (which distribution does not depend on j). It follows that the covariance matrix of u is $(\hat{\tau}_{ii'})$, where

(61) $$\tau_{ii'} = \sigma_{ii'} + \delta_{ii'}K^{-1}\sigma_e^2,$$

and $\delta_{ii'} = 1$ if $i = i'$, 0 if $i \neq i'$. An estimate of $\tau_{ii'}$ is the sample covariance of the ith row of cell means $\{y_{ij\cdot}\}$ with the i'-th row,

(62) $$\hat{\tau}_{ii'} = (J-1)^{-1}\sum_j (y_{ij\cdot} - y_{i\cdot\cdot})(y_{i'j\cdot} - y_{i'\cdot\cdot}),$$

and hence if $K > 1$, an estimate of $\sigma_{ii'}$ is

(63) $$\hat{\sigma}_{ii'} = \hat{\tau}_{ii'} - \delta_{ii'}K^{-1}\hat{\sigma}_e^2.$$

We remark that if we estimate σ_B^2 and σ_{AB}^2 by substituting the estimates (63) in (20) and (21), we get the same estimates as before in (57) and (58).

6. Distribution theory under the normality assumption. Under the normality assumption $(\mathfrak{N})$ of Section 3, the four sums of squares $(SS)_A$, $(SS)_B$, $(SS)_{AB}$, $(SS)_e$ are pairwise independent, except for the pair $(SS)_B$, $(SS)_{AB}$. We shall prove this for the pair $(SS)_A$, $(SS)_{AB}$; the independence of the other pairs may be verified similarly.

Let us write

(64) $$(SS)_A = JK\sum_{i'} L_{i'}^2,$$

(65) $$(SS)_{AB} = K\sum_i \sum_j L_{ij}^2,$$

where

(66) $$L_{i'} = A_{i'} + B_{i'},$$

(67) $$L_{ij} = A_{ij} + B_{ij},$$

(68) $$A_{i'} = \alpha_{i'} + c_{i'\cdot},$$

(69) $$B_{i'} = e_{i'\cdot\cdot} - e_{\cdot\cdot\cdot},$$

(70) $$A_{ij} = c_{ij} - c_{i\cdot},$$

(71) $$B_{ij} = e_{ij\cdot} - e_{i\cdot\cdot} - e_{\cdot j\cdot} + e_{\cdot\cdot\cdot}.$$

Then it suffices because of the joint normality of the set $\{L_{i'}, L_{ij}\}$ to prove $\text{cov}(L_{i'}, L_{ij}) = 0$ for all i', i, j. Now, any B just defined is independent of any A because of our assumption that the set $\{e_{ijk}\}$ is independent of the set $\{m_{ij}\}$. Furthermore, $B_{i'}$ and B_{ij} are orthogonal by the familiar Model I theory. Hence, it remains only to show $\text{cov}(A_{i'}, A_{ij}) = 0$:

$$(72)\quad \text{cov}\,(A_{i'}, A_{ij}) = E[c_{i'\cdot}(c_{ij} - c_{i\cdot})] = E[J^{-1}\sum_{j'} c_{i'j'}(c_{ij} - J^{-1}\sum_{j''} c_{ij''})]$$
$$= J^{-1}\sum_{j'} E(c_{i'j'}c_{ij}) - J^{-2}\sum_{j'}\sum_{j''} E(c_{i'j'}c_{ij''})$$
$$= J^{-1}E(c_{i'}c_i) - J^{-1}E(c_{i'}c_i) = 0,$$

since $E(c_{ij}c_{i'j'}) = \delta_{jj'}E(c_ic_{i'})$.

The above proof shows also that $\hat{\alpha}_i$ is statistically independent of $(SS)_{AB.i}$, since $\hat{\alpha}_i = L_i$ and $(SS)_{AB.i} = K\sum_j L_{ij}^2$.

From (55) it follows that $(SS)_e$ is distributed as σ_e^2 times χ^2 with $IJ(K-1)$ d.f. To see that $(SS)_B$ is distributed as $E(MS)_B$ times χ^2 with $J-1$ *d.f.*, write $f_j = b_j + e_{\cdot j\cdot}$ in (47), so that $(SS)_B = IK\sum_j(f_j - f_\cdot)^2$, where the set $\{f_j\}$ are independently $N(0, \sigma_f^2)$ (normal with mean 0 and variance σ_f^2) with $\sigma_f^2 = \sigma_B^2 + I^{-1}K^{-1}\sigma_e^2$, and hence $(SS)_B$ is $IK\sigma_f^2$ times χ^2 with $J-1$ d.f. Similarly, putting $g_j = c_{ij} + e_{ij\cdot} - e_{\cdot j\cdot}$ in (50), we find $(SS)_{AB.i}$ is $E(MS)_{AB.i}$ times χ^2 with $J-1$ d.f. It may be shown that for $I > 2$, $(SS)_A$ and $(SS)_{AB}$ are not, in general, distributed as a constant times noncentral (which includes central) χ^2. However, under the hypothesis H_{AB} that $\sigma_{AB}^2 = 0$, all $c_{ij} = 0$, so $(SS)_{AB}$ becomes simply

$$(73)\qquad K\sum_i\sum_j(e_{ij\cdot} - e_{i\cdot\cdot} - e_{\cdot j\cdot} + e_{\cdot\cdot\cdot})^2,$$

which is known from Model I theory to be distributed as σ_e^2 times χ^2 with $(I-1)(J-1)$ d.f.

The obvious consequence of our assumptions, that the J columns of cell means $\{y_{ij\cdot}\}$ are independently distributed like an I-variate normal vector with means $\mu_1, \cdots, \mu_I$ and covariance matrix $(\tau_{ii'})$ given by (61), we shall utilize in the next section.

7. Tests and confidence intervals. Suppose first that $K > 1$. Then the χ^2-distribution of $(SS)_e/\sigma_e^2$ affords confidence intervals for σ_e^2 in the usual way.

Since the quotient of $(MS)_B/(IK\,\sigma_B^2 + \sigma_e^2)$ by $(MS)_e/\sigma_e^2$ has the F-distribution with $J-1$ and $IJ(K-1)$ d.f., confidence intervals for σ_B^2/σ_e^2 are available as well as tests of the hypothesis that $\sigma_B^2 = 0$, or, more generally, that $\sigma_B^2/\sigma_e^2 \leqq c$, a given constant. The test at the α level of significance consists of rejecting the hypothesis if and only if $(MS)_B/(MS)_e \geqq (IKc+1)F_\alpha$, where F_α is the upper α point of the F-distribution. The power of the test can be expressed in terms of the (central) F-distribution.

The hypothesis $H_{AB}: \sigma_{AB}^2 = 0$ may be tested with the statistic $(MS)_{AB}/(MS)_e$, which, under H_{AB}, has the F-distribution with $(I-1)(J-1)$ and $IJ(K-1)$ d.f. Since this statistic is distributed as the quotient of a linear combination (with coefficients in general unequal) of independent χ^2 variables by another independent χ^2 variable, the power of the test is not expressible in terms of the noncentral F-distribution, but it could be approximated by use of a central F-distribution by using methods of Box [2].

We now drop the restriction $K > 1$. Even through $(MS)_A$ and $(MS)_{AB}$ are

statistically independent and under the hypothesis H_A: all $\alpha_i = 0$ have the same expected values, their quotient does not, in general, have the F-distribution under H_A. A test of H_A based on Hotelling's T^2 statistic is given in the next paragraph. However, confidence intervals for a particular α_i, a particular μ_i, or a particular difference $\alpha_i - \alpha_{i'}$ (none of these selected according to the outcome of the experiment) can be based on the t-distribution with $J - 1$ d.f. of the respective quotients

$$J^{1/2}K^{1/2}(\hat{\alpha}_i - \alpha_i)/(MS)_{AB,i}^{1/2}, \tag{74}$$

$$J^{1/2}(y_{i\cdot\cdot} - \mu_i)/\hat{\tau}_{ii}^{1/2}, \tag{75}$$

$$[(\hat{\alpha}_i - \hat{\alpha}_{i'}) - (\alpha_i - \alpha_{i'})]/S_{ii'}^{1/2}, \tag{76}$$

where the denominators are defined by (34), (62), and (60).

We assume now that $J \geqq I$. To calculate Hotelling's T^2 statistic for H_A, and, in case we find it significant, to make multiple comparisons, we construct a rectangular table with $R = I - 1$ rows and J columns, the entry in the rth row and jth column being

$$d_{rj} = y_{rj\cdot} - y_{Ij\cdot}, \tag{77}$$

and we compute the R means $\{d_{r\cdot}\}$ and the $\frac{1}{2}R(R + 1)$ sums of products (which, divided by $J(J - 1)$, are estimates of the covariances of the $\{d_{r\cdot}\}$)

$$a_{rr'} = \sum_j (d_{rj} - d_{r\cdot})(d_{r'j} - d_{r'\cdot}) = \sum_j d_{rj}d_{r'j} - Jd_{r\cdot}d_{r'\cdot}. \tag{78}$$

The T^2 statistic is (except for a constant factor)

$$F = J(J - I + 1)(I - 1)^{-1}Q, \tag{79}$$

where Q is the quadratic form

$$Q = \sum_r \sum_{r'} a^{rr'} d_{r\cdot}d_{r'\cdot}, \tag{80}$$

and $(a^{rr'})$ is the matrix inverse to $(a_{rr'})$. It is not necessary actually to compute the inverse matrix, since Q may be written in a form given by Rao [10] in terms of two determinants of order R calculated from $(a_{rr'})$,

$$Q = \frac{|a_{rr'} + d_{r\cdot}d_{r'\cdot}|}{|a_{rr'}|} - 1. \tag{81}$$

The statistic F in (79) has under H_A the F-distribution with $I - 1$ and $J - I + 1$ d.f., so that if F'_α denotes the upper α point of the F-distribution with these numbers of d.f., H_A is rejected at the α level of significance if and only if $F > F'_\alpha$.

The above form of the T^2 test appears to lack symmetry, since the Ith row plays a distinguished role. It is easy to see that if instead of the $\{d_{r\cdot}\}$, any basis is used for the $(I - 1)$-dimensional space spanned by the differences $\{y_{i\cdot\cdot} - y_{i'\cdot\cdot}\}$, the same test would be obtained. A symmetric form of Q (and of the noncentrality parameter δ^2 below) was given by Hsu [7], but this form would involve more numerical calculation.

The power of the T^2 test of H_A may be expressed [7] in terms of the noncentral F-distribution: The statistic (79) is distributed as noncentral F with $I - 1$ and $J - I + 1$ d.f. and noncentrality parameter δ^2, whose value will be given below, that is, as

$$(82) \qquad (I - 1)^{-1}(J - I + 1)\left[(x_1 + \delta)^2 + \sum_{r=2}^{I-1} x_r^2\right]\left(\sum_{r=I}^{J} x_r^2\right)^{-1},$$

where the $\{x_r\}$ are independently $N(0, 1)$. The noncentrality parameter δ^2 has the value

$$(83) \qquad \delta^2 = \sum_r \sum_{r'} \alpha^{rr'} \delta_r \delta_{r'} ,$$

where $(\alpha^{rr'})$ is the matrix inverse to that with elements

$$(84) \qquad \alpha_{rr'} = \text{cov}(d_{r\cdot}, d_{r'\cdot}) = J^{-1}(\tau_{rr'} - \tau_{rI} - \tau_{r'I} + \tau_{II}),$$

and

$$(85) \qquad \delta_r = \alpha_r - \alpha_I = \mu_r - \mu_I .$$

In his paper in 1931 on the T^2 test, Hotelling [6] gave an associated confidence ellipsoid. Recently, the writer [11] published a method of multiple comparison derived from the confidence ellipsoid associated with the F-test for equality of means in Model I. The same method, when based on Hotelling's confidence ellipsoid, tells us the following: Let θ be any contrast among the $\{\alpha_i\}$ or $\{\mu_i\}$, $\theta = \sum_i h_i\alpha_i = \sum_i h_i\mu_i$, where $\{h_i\}$ is any set of known constants satisfying $\sum_i h_i = 0$. Let $\hat{\theta}$ be the estimate $\hat{\theta} = \sum_i h_i y_{i\cdot\cdot} = \sum_r h_r d_{r\cdot}$, so that its variance $\sigma_{\hat{\theta}}^2 = \sum_r \sum_{r'} h_r h_{r'} \alpha_{rr'}$ is estimated by

$$(86) \qquad \hat{\sigma}_{\hat{\theta}}^2 = J^{-1}(J - 1)^{-1} \sum_r \sum_{r'} h_r h_{r'} a_{rr'} .$$

Then, for the totality of contrasts $\{\theta\}$, the probability is $1 - \alpha$ that they simultaneously satisfy

$$(87) \qquad \hat{\theta} - S\hat{\sigma}_{\hat{\theta}} \leqq \theta \leqq \hat{\theta} + S\hat{\sigma}_{\hat{\theta}} ,$$

where the constant S is calculated from F'_α, the upper α point of the F-distribution with $I - 1$ and $J - I + 1$ d.f., by

$$(88) \qquad S^2 = (I - 1)(J - 1)(J - I + 1)^{-1}F'_\alpha .$$

Whenever the T^2 test rejects H_A at significance level α, there will exist contrasts θ for which the interval (87) does not cover zero (and conversely). However, it may occasionally happen in applications that none of the contrasts thus found to be "significantly different from zero" is of any practical interest.

An interesting interpretation of the quantities $\hat{\sigma}_{\hat{\theta}}$ needed in (87), which yields an alternative way of calculating them not requiring calculation of the $\{a_{rr'}\}$ or $\{d_{rj}\}$, is the following[3]: Let $\hat{\theta}_j$ be the estimate of θ from the jth column, $\hat{\theta}_j = \sum_i h_i y_{ij\cdot}$, so $\hat{\theta} = \hat{\theta}_{\cdot}$; then

[3] Pointed out to me by Professor J. W. Tukey.

$$\hat{\sigma}_{\hat{\theta}}^2 = J^{-1}(J-1)^{-1} \sum_j (\hat{\theta}_j - \hat{\theta})^2. \tag{89}$$

However, if the calculations for the T^2 test of H_A have already been made, use of (87) may be faster.

8. Concluding remarks. The T^2 test of H_A and the associated method of multiple comparison are valid under less restrictive assumptions about the errors $\{e_{ijk}\}$. For instance, it would be sufficient that they be independently $N(0, \sigma_i^2)$, or, more generally, that the J vectors with components $e_{1j\cdot}, \cdots, e_{Ij\cdot}$ be independently distributed like a normal random vector with zero means and an arbitrary covariance matrix. The test and comparison method should be fairly insensitive to violation of the normality assumption ($\mathfrak{N}$), from consideration of the asymptotic distribution when $J \to \infty$.

A common practice in the analysis of variance is to employ as statistic to test a hypothesis the quotient of two independent mean squares whose expected values are equal under the hypothesis, and to refer this statistic to the F-tables with the numbers of d.f. equal to the ranks of the quadratic forms in the mean squares. According to this practice, $(MS)_A/(MS)_{AB}$ would be treated as though it had the F-distribution with $I - 1$ and $(I - 1)(J - 1)$ d.f. under H_A. A justification of this would be welcomed by the practitioner, because the computations are simpler and more familiar than those with Hotelling's T^2, but until numerical investigations are made which indicate the errors involved are tolerable, the practice should be suspect in the present case. The exact distribution of the statistic under H_A depends on unknown parameters. The distribution has been treated by McCarthy [8], but in a form that does not seem useful for $I > 3$. Some general theory for the distribution of ratios of statistically independent quadratic forms in jointly normal variables has recently been given by Box [2], and the above distribution falls under an application he made to another problem ([3] pp. 489–490), where he approximates it by an F-distribution with reduced numbers of d.f. However, these numbers of d.f. would depend on the covariance matrix (τ_{ij}) whose elements are defined by (61), and if we were to estimate these numbers from the data—with somewhat questionable effects on the resulting test and multiple comparisons method—it would require computation of the whole estimated covariance matrix $(\hat{\tau}_{ij})$ defined by (62). The amount of numerical work involved would then be comparable to that for the above exact methods utilizing the T^2 statistic.

An interesting practical conclusion from the present model is that the number J of levels of the Model II factor should be at least a few more than the number I of levels of the Model I factor, since the F statistic for the T^2 test has $J - I + 1$ d.f. in the denominator.

The writer acknowledges his inspiration from a paper by Tukey [13] in which the expected mean squares in the mixed model fall out as limiting cases of those obtained in sampling a finite model similar to the above with sampling of both rows and columns, as the population number of columns becomes infinite, and

with the population number of rows equal to the sample number. The effect of sampling the rows is to make all permutations of the rows equally probable and thus impose the symmetry condition ($\mathcal{S}$). However, this does not affect the expected mean squares we derived for A, B, AB, and e, since the formulas for the corresponding sums of squares are invariant under permutation of the rows. Wilk and Kempthorne [14] have recently calculated expected mean squares for a model somewhat resembling Tukey's, closer to the above in that only columns are sampled, but differing more in that the error term is generated solely by the actual randomization used to assign the "treatment combinations" to experimental units from a finite population, with the consequent introduction of treatment-unit interactions: If these are neglected the expected mean squares of Wilk and Kempthorne agree with Tukey's.

A multivariate normal model[4] for randomized blocks was studied by McCarthy [8] as an approximation to Neyman's [9] more realistic model reflecting the randomization actually used in the assignment of the varieties to the plots in each block. A test, implicitly assuming such a multivariate normal model for randomized blocks, and employing Hotelling's T^2 was recently proposed by Graybill [5]. A multivariate model for the analysis of variance was also considered by Box [1] in a different situation where the condition ($\mathcal{S}$) was tenable, and he included among other tests one of ($\mathcal{S}$). The application of Hotelling's T^2 statistic to test the equality of the components of the vector of means in samples from a multivariate normal population is due to Hsu [7].

REFERENCES

[1] G. E. P. Box, "Problems in the analysis of growth and wear curves," *Biometrics*, Vol. 6 (1950), pp. 362–389.

[2] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302.

[3] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and of correlation between errors in the two-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484–498.

[4] C. Eisenhart, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1–21.

[5] F. Graybill, "Variance heterogeneity in a randomized block design," *Biometrics*, Vol. 10 (1954), pp. 516–520.

[6] H. Hotelling, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[7] P. L. Hsu, "Notes on Hotelling's generalized T," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 231–243.

---

[4] When I discussed my results with Dr. Jerome Cornfield, he informed me that he and Dr. Max Halperin had also considered a multivariate model for the present problem and had been led to the T^2 test. Professor J. L. Hodges, Jr., formulated the above multivariate model before I did. A model equivalent to the above under the assumptions ($\mathfrak{N}$) and ($\mathcal{S}$) was proposed earlier by Mr. Leon Herbach in an unpublished paper, in which he derived the expectations and distributions of the usual mean squares and tests of the usual hypotheses.

[8] M. D. McCarthy, "On the application of the z-test to randomized blocks," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 337–359.

[9] J. Neyman, "Statistical problems in agricultural experimentation," *J. Roy. Stat. Soc., Suppl.*, Vol. 2 (1935), pp. 107–180.

[10] C. R. Rao, "Tests of significance in multivariate analysis," *Biometrika*, Vol. 35 (1948), pp. 58–79.

[11] H. Scheffé, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40 (1953), pp. 87–104.

[12] H. Scheffé, "Statistical methods for evaluation of several sets of constants and several sources of variability," *Chem. Eng. Progress*, Vol. 50 (1954), pp. 200–205.

[13] J. W. Tukey, "Interaction in a row-by-column design," Statistical Research Group, Princeton University, Memorandum Report 18 (1949).

[14] M. B. Wilk and O. Kempthorne, "Fixed, mixed, and random models in the analysis of variance," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1144–1167.

KEEPING MOMENT-LIKE SAMPLING COMPUTATIONS SIMPLE[1]

JOHN W. TUKEY

Princeton University

1. Summary. This is an attempt to present as simply as possible the best tools we know today for keeping computations simple when dealing with samples from general populations. Such computations seem inevitably to be made in terms of quantities related to moments. We develop here the formal structure and inter-relations of the two systems of multi-index quantities which seem today to be best adapted to statistical use. The occurrence of two systems is, at least in part, related to the appearance in statistical problems of both multiplication and addition of independent variables. Hence the existence of two systems, whose limiting cases are moments (about a fixed point) and cumulants (or semiinvariants).

We present interconversion formulas, developing definitions and proving the pairing formulas without reference to any infinite populations, and sparing the use of combinatorial techniques as much as we are able. A few multiplication formulas are given, but for a more complete list the reader is referred to Wishart [10]. It is hoped that this paper can be read on its own, with some reference to applications of these techniques to elementary examples [6] and to the sampling properties of estimated variance components in the analysis of variance [7], [8], [9] as motivation.

The author's best thanks go to N. R. Goodman for the checking of certain calculations.

2. Introduction. The history of "moments of moments", still the only way we know to attack general sampling distributions, has been long and complicated. Its outstanding feature has been the cutting away of pages and pages of algebra by the introduction of new and sharper tools. The forging of these tools has depended more and more on combinatorial ideas, and while the use of the tools has become simpler, their understanding has become apparently more and more complex. It is the purpose of this paper to show how we can keep almost everything quite simple and still use what today seem to be the best tools. The only useful aspects which we cannot completely handle simply are the actual calculation of certain multiplication formulas, an extensive table of which has been provided by Wishart [10].

The two systems which we shall discuss correspond to moments about a fixed point and to cumulants (semiinvariants). They do not correspond to moments

Received April 6, 1955.

[1] Prepared in connection with research sponsored by the Office of Naval Research. Based on a part of Statistical Research Group, Princeton University, Memorandum Report 45 "Finite Sampling Simplified," written while the author was a Fellow of the John Simon Guggenheim Foundation.

about the mean, except insofar as the second and third moments about the mean, happen to be cumulants. From the point of view of practical use, either numerical or algebraic, I am convinced that higher moments *about the mean* are a vermiform appendix of statistical evolution—an evolutionary remnant which will inevitably disappear, though all too slowly.

The two systems involve not a single index, but a set of one or more indices. At first sight this may seem redundant and wasteful, since there are relations which could be used to eliminate the multiple index symbols. But these relations involve sizes of samples or populations, and a great part of the simplicity of the use of the multiple index systems arises from a great reduction in the appearance of these sizes.

We shall work entirely in terms of finite samples or populations, treating infinite populations as special limiting cases. Contrary to the usual view, this does not make matters more complex.

3. The two systems. The first system of polynomial symmetric functions of n numbers $x_1, x_2, \cdots, x_n$ which we shall use are the symmetric means (the mean power products, in combinatorial terminology) which we will denote by angle brackets, as $\langle 3\rangle$, $\langle 134\rangle$, etc. and will refer to as symmetric means or brackets. They are defined as the means of products of powers of *different* x_i's, so that, for example,

$$\langle ab\rangle = \frac{\sum^{\neq} x_i^a x_j^b}{n(n-1)}$$

where the sum is over the $n(n-1)$ pairs (i, j) with $i \neq j$. The numerators are a kind of symmetric function of very respectable antiquity, the only modern features being (i) the division by the number of cases to give a mean, which seems natural to the statistician but perhaps not to the combinatorialist, and (ii) the use of multiple subscripts, which the combinatorialist has always done but which the statistician has seemed to resist. This resistance seems to have been due to a feeling that only the simple moments

$$\langle a\rangle = \frac{\sum x_i^a}{n}$$

could be easily calculated from numerical data, and that hence all formulas should be written in terms of moments. This position is tempting rather than irrefutable, and the simplicity of formulas involving the multiple subscripts shows its deficiencies. It is easy to continue the numerical calculation, once the moments are at hand, and find all the symmetric functions of either system. For all of weight ≤ 4 a simple computing form has been presented in [6].

The second, and even more important, system of polynomial symmetric functions is most simply defined in terms of the first system through linear formulas like these

$$k_2 = (2) = \langle 2\rangle - \langle 11\rangle,$$

$$k_{113} = (113) = \langle 113\rangle - 3\langle 1112\rangle + 2\langle 11111\rangle.$$

We shall obtain the general law of formation and supply a compact table of formulas up to weight 8. We shall usually refer to these as "polykays" for the sake of a short, simple term. (While the construction of this term, "kay" for "k" and "poly" for the multiple subscript, seems somewhat revolting to some colleagues, the use of "generalized k-statistics" seems too unhandy to me. In due course, perhaps, someone will find a good, short, terminology.) The polykays with but a single subscript are, of course, Fisher's famous k-statistics, whose introduction was perhaps the largest step in clearing unnecessary algebra out of this field. The multiple index analogs were introduced by Dressel [3] in 1940 in a combinatorial paper which seems to have escaped notice at the time of its appearance. They were introduced independently by the author in 1950 [6] as practical working tools.

4. Random pairing, additive and multiplicative. Let us next consider two sets of n numbers, $x_1^*, x_2^*, \cdots, x_n^*$ and $x_1^{**}, x_2^{**}, \cdots, x_n^{**}$, whose brackets and polykays we shall similarly distinguish with asterisks, as, for example, $\langle 2\rangle^*$ and $\langle 2\rangle^{**}$. We shall be concerned with the results of *pairing* these two sets randomly, more specifically with the results of forming some function of each of the pairs

$$[x_1^*, x_{\pi(1)}^{**}], [x_2^*, x_{\pi(2)}^{**}], \cdots, [x_n^*, x_{\pi(n)}^{**}]$$

where $\pi(1), \pi(2), \cdots, \pi(n)$ is a permutation of the integers $1, 2, \cdots, n$, and where we shall eventually wish to average over all permutations.

The simplest pairing operation is multiplicative pairing, where $x_i = x_i^* x_{\pi(i)}^{**}$. Let us calculate a moment of the resulting x_i, say $\langle a\rangle$, and then average over all pairings [permutations]. We have

$$\langle a\rangle = \frac{\sum x_i^a}{n} = \frac{\sum (x_i^*)^a (x_{\pi(i)}^{**})^a}{n}$$

and when we average, the product of x_i^* and x_j^{**} appears equally often for *all pairs* i and j, so that

$$\text{aver } \{\langle a\rangle\} = \frac{\sum (x_i^*)^a \cdot \sum (x_j^{**})^a}{n^2} = \langle a\rangle^* \langle a\rangle^{**}$$

where we have written "aver" for the *av*erage over *r*andom pairing, as we shall continue to do, and where the denominator of n^2 is easily justified as equal to the number of terms in the numerator when expanded (an average of a mean is again a mean).

The same argument applies to $\langle ab \cdots e\rangle$, as we see if we write

$$y_{ij\cdots m} = x_i^a x_j^b \cdots x_m^e$$

$$y_{ij\cdots m}^* = (x_i^*)^a (x_j^*)^b \cdots (x_m^*)^e$$

$$y_{ij\cdots m}^{**} = (x_i^{**})^a (x_j^{**})^b \cdots (x_m^{**})^e$$

and observe that

$$y_{ij\cdots m} = y^*_{ij\cdots m} y^{**}_{\pi(i),\pi(j),\cdots,\pi(m)} .$$

Thus we have

$$\text{aver } \{\langle ab \cdots e\rangle\} = \langle ab \cdots e\rangle^* \langle ab \cdots e\rangle^{**}.$$

The brackets are ideally suited to multiplicative pairing.

The only other random pairing which we know how to handle at all well statistically is the additive one, where

$$x_i = x_i^* + x^{**}_{\pi(i)} .$$

Because of the many statistical problems where additive pairing is taken as the first approximation to reality (classical analysis of variance models, error propagation, etc.), this is the most important case.

The one-index (or as we would say in Section 11, one-part) brackets behave in a manageable, but not simple way under additive random pairing. We have, for example,

$$\begin{aligned}\langle\langle 3\rangle\rangle = \text{aver } \{\langle 3\rangle\} &= \text{aver } \frac{\sum (x_i^* + x^{**}_{\pi(i)})^3}{n} \\ &= \text{aver } \frac{\sum (x_j^*)^3}{n} + 3 \text{ aver } \frac{\sum (x_i^*)^2 (x^{**}_{\pi(i)})}{n} + 3 \text{ aver } \frac{\sum (x_i^*)(x^{**}_{\pi(i)})^2}{n} \\ &\qquad + \text{aver } \frac{\sum (x^{**}_{\pi(i)})^3}{n} \\ &= \langle 3\rangle^* + 3\langle 2\rangle^*\langle 1\rangle^{**} + 3\langle 1\rangle^*\langle 2\rangle^{**} + \langle 3\rangle^{**}\end{aligned}$$

and, in general,

$$\begin{aligned}\langle\langle j\rangle\rangle = \text{aver } \{\langle j\rangle\} &= \langle j\rangle^* + \binom{j}{1}\langle j-1\rangle^*\langle 1\rangle^{**} \\ &\qquad + \binom{j}{2}\langle j-2\rangle^*\langle 2\rangle^{**} + \cdots + \langle j\rangle^{**}\end{aligned}$$

where we have introduced a doubling of the brackets to indicate averaging over an *additive* random pairing.

Since the general formula is of binomial type, we can represent it in terms of generating functions. If we define

$$\begin{aligned}M_{\text{aver}}(t) &= 1 + \langle\langle 1\rangle\rangle t + \langle\langle 2\rangle\rangle \frac{t^2}{2!} + \langle\langle 3\rangle\rangle \frac{t^3}{3!} + \cdots , \\ M^*(t) &= 1 + \langle 1\rangle^* t + \langle 2\rangle^* \frac{t^2}{2!} + \langle 3\rangle^* \frac{t^3}{3!} + \cdots , \\ M^{**}(t) &= 1 + \langle 1\rangle^{**} t + \langle 2\rangle^{**} \frac{t^2}{2!} + \langle 3\rangle^{**} \frac{t^3}{3!} + \cdots ,\end{aligned}$$

then the general formula becomes

$$M_{\text{aver}}(t) = M^*(t)M^{**}(t).$$

We shall use this relation in Section 11 to obtain general expressions defining the second system of quantities in relation to the brackets.

5. The polykays. This second system will be denoted either by $k_{ab\cdots e}$ or by $(ab \cdots e)$ as may be convenient. We shall use the same double parenthesis convention, so that

$$((ab \cdots e)) = \text{aver}\,\{(ab \cdots e)\}$$

where the averaging is over *additive* random pairing.

Two examples of the second system, beyond the trivial $(1) = \langle 1\rangle$, are $(2) = \langle 2\rangle - \langle 11\rangle$ and $(12) = \langle 12\rangle - \langle 111\rangle$. Let us examine the behavior of these quantities under random pairing, using unproven, but formally reasonable, facts about the behavior of multipart brackets. We find

$$\begin{aligned}((2)) &= \langle\langle 2\rangle\rangle - \langle\langle 11\rangle\rangle = \langle 2\rangle^* + 2\langle 1\rangle^*\langle 1\rangle^{**} + \langle 2\rangle^{**} - \langle 11\rangle^* - 2\langle 1\rangle^*\langle 1\rangle^{**} - \langle 11\rangle^{**}\\ &= [\langle 2\rangle^* - \langle 11\rangle^*] + [\langle 2\rangle^{**} - \langle 11\rangle^{**}] = (2)^* + (2)^{**}\end{aligned}$$

$$\begin{aligned}((12)) &= \langle\langle 12\rangle\rangle - \langle\langle 111\rangle\rangle\\ &= \langle 12\rangle^* + \langle 2\rangle^*\langle 1\rangle^{**} + 2\langle 11\rangle^*\langle 1\rangle^{**} + \langle 1\rangle^*\langle 2\rangle^{**} + 2\langle 1\rangle^*\langle 11\rangle^{**} + \langle 12\rangle^{**}\\ &\qquad - \langle 111\rangle^* - 3\langle 11\rangle^*\langle 1\rangle^{**} - 3\langle 1\rangle^*\langle 11\rangle^{**} - \langle 111\rangle^{**}\\ &= [\langle 12\rangle^* - \langle 111\rangle^*] + [\langle 2\rangle^* - \langle 11\rangle^*]\langle 1\rangle^{**} + \langle 1\rangle^*[\langle 2\rangle^{**} - \langle 11\rangle^{**}]\\ &\qquad + [\langle 12\rangle^{**} - \langle 111\rangle^{**}]\\ &= (12)^* + (2)^*(1)^{**} + (1)^*(2)^{**} + (12)^{**}\end{aligned}$$

It should now be clear both what the pairing law is likely to be, and that we need some slick trick both to discover the definitions and to prove the result.

Since the trick the author prefers involves symbolic multiplication, we shall postpone its treatment to the last section. We announce here the pairing formula we desire and leave definitions and proofs to Sections 11 and 12. The pairing formula is illustrated by

$$\begin{aligned}\text{ave}\,\{k_{abc}\} = ((abc)) &= (abc)^* + (ab)^*(c)^{**} + (ac)^*(b)^{**} + (bc)^*(a)^{**}\\ &\qquad + (a)^*(bc)^{**} + (b)^*(ac)^{**} + (c)^*(ab)^{**} + (abc)^{**}\end{aligned}$$

In general, we separate the indices into two sets (one of which may be empty) in all possible ways, assigning one set to * and the other to **, and adding up the resulting products. We know that we have the desired definitions when the pairing formula is an identity with $((ab \cdots e))$ the same function of the $\langle\langle fg \cdots j\rangle\rangle$ as $(ab \cdots e)^*$ is of the $\langle fg \cdots j\rangle^*$ and as $(ab \cdots e)^{**}$ is of the $\langle fg \cdots j\rangle^{**}$.

One special case is worthy of notice. If all the x_i^{**} are identically equal to δ, then, (cp. [7])

$$\langle a\rangle^{**} = \delta^a; \quad \langle ab\rangle^{**} = \delta^{a+b}; \quad \langle abc\rangle^{**} = \delta^{a+b+c}; \cdots; (1)^{**} = \delta;$$

$$(a)^{**} = 0, \text{ for } a > 1; \quad 0 = (ab)^{**} = (abc)^{**} = (abcd)^{**} = \cdots.$$

So that the effect of pairing with this population, which is independent of the randomization and exactly equivalent to increasing all the x_i^* by δ, is given by such relations as

$$(1) = (1)^* + \delta, \quad (2) = (2), \quad (3) = (3),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\begin{aligned}
(11) &= (11)^* + 2\delta(1)^* + \delta^2, \\
(12) &= (12)^* + \delta(2)^*, \\
(111) &= (111)^* + 3\delta(111)^* + 3\delta^2(1)^* + \delta^3 \\
(112) &= (112)^* + 2\delta(12)^* + \delta^2(2)^*, \\
(22) &= (22)^*.
\end{aligned}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

Thus the effects of increasing all the x_i^* by δ are easily found. We notice for future use that the highest power of δ is the number of 1's in the polykay; which appears with a coefficient found by dropping the 1's from within the parentheses.

6. Inheritance and representation. If $x_1, x_2, \cdots, x_n$ is a sample from x_1', $x_2', \cdots, x_N'$, and if "ave" means a simple average over all samples, then symmetry implies that

$$\text{ave}\,\{\langle ab \cdots e\rangle\} = \frac{\sum \text{aver}\,\{x_i^a x_j^b \cdots x_m^e\}}{\text{number of terms above}} = \frac{\sum^{\neq} (x_i')^a (x_j')^b \cdots (x_m')^e}{\text{number of terms above}}$$

$$= \langle ab \cdots e\rangle'.$$

Since the polykays are expressible linearly in the brackets with integer coefficients the same relation

$$\text{ave}\,\{k_{ab\cdots e}\} = k'_{ab\cdots e}$$

must hold for polykays. We refer to this as inheritance on the average. Some would prefer to say unbiasedness (but there are now so many kinds of unbiasedness!).

Since any polynomial symmetric function can be expressed linearly in terms of the brackets, and since, as we shall see, brackets and polykays can be expressed linearly in terms of one another, every polynomial symmetric function can also be expressed linearly in terms of polykays.

7. Unit parts. Each index "1" which appears in a bracket or a polykay will be called a unit part. These indices play an especially simple role. If we have a linear identity connecting polykays and brackets, we can obtain a new one

by adding unit parts to all the terms. Thus from $k_3 = (3) = \langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle$ we derive $k_{13} = (13) = \langle 13\rangle - 3\langle 112\rangle + 2\langle 1111\rangle$, $k_{113} = (113) = \langle 113\rangle - 3\langle 1112\rangle + 2\langle 11111\rangle$ and so on, while from $\langle 2\rangle = (2) + (11)$ we derive $\langle 12\rangle = (12) + (111) = k_{12} + k_{111}$, $\langle 112\rangle = (112) + (1111) = k_{112} + k_{1111}$ and so on. This fact will be proven in Section 12.

Thus it would be convenient to write these formulas in a single shorthand form, such as $(-3) = \langle -3\rangle - 3\langle -2\rangle + 2\langle -\rangle$, $\langle -2\rangle = (-2) + (-)$ where the dashes stand for the number of 1's. This number is usually different in the different appearances in one formula, but these numbers are to be chosen so as to make all terms of the same degree.

Fur further abbreviation, we may drop the dashes themselves (except in $(-)$ and $\langle -\rangle$ where they seem helpful). If we do this, then we can give a single table which presents the coefficients for all the identities connecting brackets and polykays through degree 8. This is done in a compact form (pioneered by David and Kendall [2]) in Table 1, where brackets are expressed in terms of polykays by the coefficients below and on the main diagonal, while polykays are expressed in terms of brackets by the coefficients above and on the main diagonal. (A less convenient table, extending through weight 12, has recently been provided by Abdel-Aty [1].)

More specifically, to express $\langle 1134\rangle$ in terms of polykays, we proceed as follows:

(a) look for $\langle 34\rangle$, which identifies a row, and start with the heavy 1 on the diagonal of that row,

(b) move along that row from the diagonal 1 toward the beginning, writing down the coefficients times the corresponding polykays, (This yields:

$$\langle 34\rangle = (34) + 3(223) + 4(33) + 3(24) + 9(222) + 18(23) + (4) + 21(22) + 5(3) + 9(2) + (-)$$

in shorthand notation.)

(c) note that $1 + 1 + 3 + 4 = 9$ and add 1's to every term to bring the degree of every term up to 9. (This yields:

$$\begin{aligned}\langle 1134\rangle = (1134) &+ 3(11223) + 4(11133) + 3(11124) + 9(111222)\\ &+ 18(111123) + (111114) + 21(1111122)\\ &+ 5(1111113) + 9(11111112) + (111111111)\end{aligned}$$

which is the desired result.)

To expand a polykay in brackets we merely interchange rows and columns, moving upward from the diagonal.

8. Computation modulo unit parts. We see easily, either from the general relations of Sections 11 and 12, or from the nature of the reduced formulas, that when a bracket with g unit parts is written in terms of polykays, only polykays with at least g unit parts appear and vice versa. It is then unequivocal if we write

TABLE I

Conversion Formulae

	(1)	(2)	(3)	(22)	(4)	(23)	(222)	(5)	(24)	(33)	(223)	(2222)	(6)	(25)	(34)	(224)	(233)	(7)	(26)	(35)	(44)	(8)	
⟨–⟩	1	−1	2	1	−6	−2	−1	24	6	4	2	1	−120	−24	−12	−6	−4	720	120	48	36	−5040	
⟨2⟩	1	1	−3	−2	12	5	3	−60	−18	−12	−7	−4	360	84	42	24	16	−2520	−480	−192	−144	20160	⟨2⟩
⟨3⟩	1	3	1		−4	−1		20	4	4	1		−120	−20	−14	−4	−4	840	120	64	48	−6720	⟨3⟩
⟨22⟩	1	2		1	−3	−3	−3	30	15	9	8	6	−270	−90	−42	−33	−21	2520	630	240	180	−25200	⟨22⟩
⟨4⟩	1	6	4	3	1			−5	−1				30	5	2	1		−210	−30	−10	−12	1680	⟨4⟩
⟨23⟩	1	4	1	3		1		−10	−4	−6	−2		120	30	24	8	10	−1260	−240	−140	−96	13440	⟨23⟩
⟨222⟩	1	3		3			1		−3		−3	−4	30	30	9	18	9	−630	−300	−90	−72	10080	⟨222⟩
⟨5⟩	1	10	10	15	5	10		1					−6	−1				42	6	2		−336	⟨5⟩
⟨24⟩	1	7	4	9	1	4	3		1				−15	−5	−3	−2		210	45	15	24	−2520	⟨24⟩
⟨33⟩	1	6	2	9		6				1			−10		−4		−1	140	10	20	16	−1680	⟨33⟩
⟨223⟩	1	5	1	7		2	3				1			−10	−3	−4	−6	210	120	60	24	−5040	⟨223⟩
⟨2222⟩	1	4		6			4					1				−3			30		9	−630	⟨2222⟩
⟨6⟩	1	15	20	45	15	60	15	6	15	10			1					−7	−1			56	⟨6⟩
⟨25⟩	1	11	10	25	5	20	15	1	5		10			1				−21	−6	−3		336	⟨25⟩
⟨34⟩	1	9	5	21	1	18	9		3	4	3				1			−35		−5	−8	560	⟨34⟩
⟨224⟩	1	8	4	16	1	8	12		2		4	3				1			−15		−6	420	⟨224⟩
⟨233⟩	1	7	2	15		8	9			1	6						1		−10	−10		560	⟨233⟩
⟨7⟩	1	21	35	105	35	210	105	21	105	70	105		7	21	35			1				−8	⟨7⟩
⟨26⟩	1	16	20	60	15	80	60	6	30	10	60	15	1	6		15	10		1			−28	⟨26⟩
⟨35⟩	1	13	11	45	5	50	45	1	15	10	45			3	5		10			1		−56	⟨35⟩
⟨44⟩	1	12	8	42	2	48	36		12	16	24	9			8	6					1	−35	⟨44⟩
⟨8⟩	1	28	56	210	70	560	420	56	420	280	840	105	28	168	280	210	280	8	28	56	35	1	⟨8⟩
	(1)	(2)	(3)	(22)	(4)	(23)	(222)	(5)	(24)	(33)	(223)	(2222)	(6)	(25)	(34)	(224)	(233)	(7)	(26)	(35)	(44)	(8)	

(To use pivot on diagonal of ones)

$O(1^g)$ for an arbitrary set of terms each of which, when expanded linearly in brackets or polykays contains at least g unit parts.

If we have an identity which is a polynomial in the polykays, as for example

$$k_1^3 - k_2k_1 + \frac{1}{n}k_3 + k_{21} \equiv \frac{1}{n^2}k_3 + \frac{3}{n}k_{21} + k_{111}$$

each term has a certain total number of unit parts. In the example, these numbers are 3, 1, 0 and 1 on the left-hand side and 0, 1 and 3 on the right-hand side. The highest total appearing on either side is the unit weight of that side. In the example, the unit weight of each side is 3. If we shift all the x_i by pairing them with a set of values all equal to δ, each term will be replaced by a number of terms involving various powers of δ up to and including a power equal to the number of unit parts. The total coefficients of each power of δ must also give an identity. Thus any identity gives rise to a number of associated identities.

If one side of the initial identity was linear in the polykays, and all obvious cancellations had been made initially (as is the case on the right-hand side of the example), then the coefficient of the highest power of δ appearing on that side after pairing would be linear in the polykays, would have no obvious cancellations, and hence would not vanish identically. The coefficient of the same power of δ on the other side, which is identically equal to this, cannot vanish, and hence

$$(\text{Unit weight on other side}) \geqq (\text{Unit weight on linear side})$$

Since any polynomial in the polykays is a symmetric polynomial in the x's, it can be written linearly in the polykays. The unit weight will not be increased in this process. In particular, a polynomial in polykays without unit part (and hence of unit weight zero) when written linearly in the polykays involves no polykay with unit part. If we know the linear representation to terms $O(1)$, we know it exactly. Similarly, the unit weight of the left-hand side in the example is 3. If we know the linear representation (the right-hand side) to $O(1^4)$, we know it exactly.

In table 1 we have both heavy and light coefficients. Only the heavy ones need to be used if we adjoin $+O(1)$. Thus, for example,

$$\begin{aligned}\langle 34\rangle &= (34) + 3(223) + O(1)\\(34) &= \langle 34\rangle - 3(223) + O(1)\\(223) &= \langle 223\rangle + O(1)\end{aligned}$$

As this example shows, the formulas are often much simplified by this process.

9. Multiplication of brackets. Both brackets and polykays are chosen so as to remove the inevitable combinatorial difficulties from as many formulas as possible. As a result, combinatorial considerations have been restricted to the formulas for multiplication. For brackets, the resulting formulas are relatively simple. Thus

$$\langle a\rangle\langle b\rangle = \frac{1}{n^2}\sum x_i^a \sum x_j^b$$
$$= \frac{1}{n^2}\left[\sum^{\neq} x_i^a x_j^b + \sum x_k^{a+b}\right]$$
$$= \frac{1}{n^2}\left[n(n-1)\langle ab\rangle + n\langle a+b\rangle\right]$$
$$= \frac{n-1}{n}\langle ab\rangle + \frac{1}{n}\langle a+b\rangle$$

in a similar way we find

$$\begin{aligned}\langle abc\rangle\langle de\rangle = {} & \frac{(n-3)(n-4)}{n(n-1)}\langle abcde\rangle + \frac{n-3}{n(n-1)}\langle a+d, bce\rangle \\ & + \frac{n-3}{n(n-1)}\langle a+e, bcd\rangle + \frac{n-3}{n(n-1)}\langle b+c, ace\rangle \\ & + \frac{n-3}{n(n-1)}\langle b+e, acd\rangle + \frac{n-3}{n(n-1)}\langle c+d, bce\rangle \\ & + \frac{n-3}{n(n-1)}\langle c+e, abd\rangle + \frac{1}{n(n-1)}\langle a+d, b+e, c\rangle \\ & + \frac{1}{n(n-1)}\langle a+d, c+e, b\rangle + \frac{1}{n(n-1)}\langle b+d, c+e, a\rangle \\ & + \frac{1}{n(n-1)}\langle b+d, a+e, c\rangle + \frac{1}{n(n-1)}\langle c+d, a+e, b\rangle \\ & + \frac{1}{n(n-1)}\langle c+d, b+e, a\rangle.\end{aligned}$$

In general, we obtain all brackets which can be obtained by matching some (including none) of the letters in one bracket with letters in the other and then replacing matched letters by their sum. The coefficient is a simple function of the number of parts in the factors *and* the result, with a simple denominator. It is often convenient to expand coefficients as integer coefficient combinations of

$$p = \frac{1}{n}, \quad q = \frac{1}{n(n-1)}, \quad r = \frac{1}{n(n-1)(n-2)},$$
$$s = \frac{1}{n(n-1)(n-2)(n-3)}, \quad \cdots.$$

These expansions are given for certain products in Table 2. With the aid of this table, multiplication of brackets is merely a matter of exerting moderate patience to be sure that you have all the terms.

10. Multiplication of polykays. The multiplication formulas for polykays are more complex. They can be obtained symbolically (cp. Wishart [10], Kendall

TABLE 2

Expansions of factors for bracket multiplication

Parts of factors	Parts in bracket whose coefficient is sought					
	1	2	3	4	5	6
1×1	p	$1 - p$	—	—	—	—
1×2	—	p	$1 - 2p$	—	—	—
1×3	—	—	p	$1 - 3p$	—	—
1×4	—	—	—	p	$1 - 4p$	—
2×2	—	q	$p - q$	$1 - 4p + 2q$	—	—
2×3	—	—	q	$p - 2q$	$1 - 6p + 6q$	—
2×4	—	—	—	q	$p - 3q$	$1 - 8p + 12q$
3×3	—	—	r	$q - r$	$p - 4q + 2r$	$1 - 9p + 18q - 6r$
3×4	—	—	—	r	$q - 2r$	$p - 6q + 6r$
4×4	—	—	—	s	$r - s$	$q - 4r + 2s$

[5]) or by direct calculation. One way to carry out direct calculation is to express each polykay in brackets, multiply out the brackets and then reconvert the resulting brackets to polykays. For example

$$\begin{aligned}
k_{12}k_2 &= (12)(2) = [\langle 12\rangle - \langle 111\rangle][\langle 2\rangle - \langle 11\rangle] \\
&= \langle 12\rangle\langle 2\rangle - \langle 111\rangle\langle 2\rangle - \langle 12\rangle\langle 11\rangle + \langle 111\rangle\langle 11\rangle \\
&= [1 - 2p]\langle 122\rangle + p\langle 23\rangle + p\langle 14\rangle - [1 - 3p]\langle 1112\rangle \\
&\quad - 3p\langle 113\rangle - [1 - 4p + 2q]\langle 1112\rangle - 2[p - q]\langle 122\rangle \\
&\quad - 2[p - q]\langle 113\rangle - 2q\langle 23\rangle + [1 - 6p + 6q]\langle 11111\rangle \\
&\quad + 6[p - 2q]\langle 1112\rangle + 6q\langle 122\rangle \\
&= [p - 2q]\langle 23\rangle + p\langle 14\rangle + [1 - 4p + 8q]\langle 122\rangle \\
&\quad + [-5p + 2q]\langle 113\rangle + [-2 + 13p - 14q]\langle 1112\rangle \\
&\quad + [1 - 6p + 6q]\langle 11111\rangle \\
&= [p - 2q][(23) + 3(122) + (113) + 4(1112) + (11111)] \\
&\quad + p[(14) + 3(122) + 4(113) + 6(1112) + (111111)] \\
&\quad + [1 - 4p + 8q][(122) + 2(1112) + (11111)] \\
&\quad + [-5p + 2q][(113) + 3(1112) + (11111)] \\
&\quad + [-2 + 13p - 14q][(1112) + (11111)] + [1 - 6p + 6q](11111) \\
&= [p - 2q](23) + p(14) + [1 + 2p + 2q](122) + O(113) - O(1112) \\
&\quad + O(11111) \\
&= \frac{n - 3}{n(n - 1)}\, k_{23} + \frac{1}{n}\, k_{14} + \frac{n + 1}{n - 1}\, k_{122}
\end{aligned}$$

Even for this case, some care in computation is advisable. Clearly direct computation should be avoided to the greatest extent possible.

In some cases, it is possible to obtain a substantial saving in computation by calculating modulo unit parts in a suitable sense. Thus in the example just given we may neglect terms $O(1^2)$. Thus we could have written the products of the brackets as

$$\begin{aligned}
&[1 - 2p]\langle 122\rangle + p\langle 23\rangle + p\langle 14\rangle - 2(p - q)\langle 122\rangle - 2q\langle 23\rangle + 6q\langle 122\rangle + O(1^2)\\
&\quad = [1 - 4p + 8q][(122) + O(1^2)] + [p - 2q][(23) + 3(122) + O(1^2)]\\
&\qquad\qquad + p[(14) + 3(122) + O(1^2)] + O(1^2),
\end{aligned}$$

which avoids a certain amount of algebra.

As a more complex example, let us take

$$\begin{aligned}
k_{22}k_{22} = (22)(22) &= (\langle 22\rangle - 2\langle 112\rangle + \langle 1111\rangle)(\langle 22 - 2\langle 112\rangle + \langle 1111\rangle)\\
&= \langle 22\rangle\langle 22\rangle - 4\langle 22\rangle\langle 112\rangle + 4\langle 112\rangle\langle 112\rangle - 4\langle 112\rangle\langle 1111\rangle\\
&\qquad + 2\langle 22\rangle\langle 1111\rangle + \langle 1111\rangle\langle 1111\rangle
\end{aligned}$$

where

$$\begin{aligned}
\langle 22\rangle\langle 22\rangle &= [1 - 4p + 2q]\langle 2222\rangle + 4[p - q]\langle 224\rangle + 2q\langle 44\rangle,\\
\langle 22\rangle\langle 112\rangle &= 2q\langle 233\rangle + O(1),\\
\langle 112\rangle\langle 112\rangle &= 2[q - r]\langle 2222\rangle + 2r\langle 224\rangle + 4r\langle 233\rangle + O(1),\\
\langle 22\rangle\langle 1111\rangle &= O(1),\\
\langle 112\rangle\langle 1111\rangle &= O(1),\\
\langle 1111\rangle\langle 1111\rangle &= 24s\langle 2222\rangle + O(1),
\end{aligned}$$

so that

$$\begin{aligned}
k_{22}k_{22} = [1 - 4p + 10q - 8r + 24s]\langle 2222\rangle + [p - 4q + 8r]\langle 224\rangle\\
+ [-8q + 16r]\langle 233\rangle + 2q\langle 44\rangle + O(1).
\end{aligned}$$

But

$$\begin{aligned}
\langle 44\rangle &= (44) + 6(224) + 9(2222) + O(1),\\
\langle 224\rangle &= (224) + 3(2222) + O(1),\\
\langle 233\rangle &= (233) + O(1),\\
\langle 2222\rangle &= (2222) + O(1),
\end{aligned}$$

so that

$$k_{22}k_{22} = [1 + 8p + 16q + 16r + 24s](2222) + [4p + 8q + 8r](224) + [-8q + 16r](233) + 2q(44)$$

$$= \left[1 + \frac{8}{n-2} + \frac{24}{n(n-1)(n-2)(n-3)}\right] k_{2222} + \frac{4}{n-2} k_{224} - \frac{8(n-4)}{n(n-1)(n-2)} k_{233} + \frac{2}{n(n-1)} k_{44} .$$

This result agrees with $k_{22}k_{22}$ as obtained from Wishart's formulas in the form

$$k_{22}k_{22} = \left(\frac{n-1}{n+1} k_2^2 - \frac{n-1}{n(n+1)} k_4\right)^2$$

$$= \left(\frac{n-1}{n+1}\right)^2 k_2^4 - 2 \frac{(n-1)^2}{n(n+1)^2} k_2^2 k_4 + \frac{(n-1)^2}{n^2(n+1)^2} k_4^2$$

and thus provides an additional check on Wishart's result.

Many of Wishart's formulas were independently obtained by the writer before their publication by Wishart. In most cases agreement was good, and in the others the writer's algebra proved at fault.

A few of the simplest are now given for easy reference:

$$k_2^2 = k_{22} + \frac{1}{n} k_4 + \frac{2}{n-1} k_{22},$$

$$k_1 k_a = k_{1a} + \frac{1}{n} k_{a+1},$$

$$k_1 k_{ab} = k_{ab1} + \frac{1}{n} k_{a+1,b} + \frac{1}{n} k_{b+1,a},$$

$$k_2 k_3 = k_{23} + \frac{1}{n} k_5 + \frac{6}{n-1} k_{23}.$$

For others the reader is referred to Wishart [10]. If products of weights greater than 8 are ever needed, it is very probably that terms in $1/n$, or perhaps through $1/n^2$ will suffice. In such cases, an extension of Table 2, neglecting r, s, t, $\cdots$ (and q if terms in $1/n$ will suffice) forms the basis of a simple method of calculation.

11. The o-multiplication and one-part k's. We know (cp. Section 4) that the generating functions $\{\langle\langle a\rangle\rangle = \text{aver } \langle a\rangle\}$, $\{\langle a\rangle^*\}$ and $\{\langle a\rangle^{**}\}$ satisfy the relation

$$M_{\text{aver}}(t) = M^*(t)M^{**}(t)$$

where the $\langle\langle a\rangle\rangle$ are the averages over all pairings of the $\langle a\rangle$ which are defined for all pairings of the sets defining the $\langle a\rangle^*$ and $\langle a\rangle^{**}$. To obtain the one-part k's without reference to the theory of infinite populations, and to prepare the ground work for the introduction of the multipart k's, we introduce a symbolic

multiplication among the following quantities: real numbers, all $\langle\langle ab \cdots e\rangle\rangle$, all $\langle ab \cdots e\rangle$, all $\langle ab \cdots e\rangle^*$, all $\langle ab \cdots e\rangle^{**}$, the integer powers of an indeterminate t, and all linear combinations of the above. This multiplication is written "o" and is defined to satisfy:

(1) *except* when a bracket is multiplied by a bracket of the same family, o-multiplication of the elementary quantities is ordinary multiplication,

(2) o-multiplication is distributive with respect to addition, subtraction and multiplication by real numbers,

(3) o-multiplication of brackets from the same family is accomplished by combining indices, as in $\langle 23\rangle \, o \, \langle 14\rangle = \langle 2314\rangle = \langle 1234\rangle$. As examples of rule 3 we have

$$\langle 11\rangle^* \, o \, \langle 34\rangle^* = \langle 1134\rangle^*,$$

$$\langle 2\rangle^{**} \, o \, (\langle 2\rangle^{**} - \langle 11\rangle^{**}) = \langle 22\rangle^{**} - \langle 112\rangle^{**},$$

where rule 2 was used in the latter case, while rule 1 shows that

$$\langle\langle 2\rangle\rangle \, o \, \langle 1\rangle^* = \langle\langle 2\rangle\rangle\langle 1\rangle^*,$$

$$\langle 3\rangle \, o \, \langle 24\rangle^{**} = \langle 3\rangle\langle 24\rangle^{**}.$$

In terms of this symbolic multiplication we have a commutative ring with formal power series in t. We can form formal o-exponentials and o-logarithms of appropriate expressions, and these functions will have the usual formal properties. Thus

$$o\text{-exp}\,(tX) = 1 + tX + \frac{t^2}{2}(X \, o \, X) + \frac{t^3}{6}(X \, o \, X \, o \, X) + \frac{t^4}{24}(X \, o \, X \, o \, X \, o \, X)t \cdots,$$

$$o\text{-log}\,(1 + tX) = tX - \frac{t^2}{2}(X \, o \, X) + \frac{t^3}{3}(X \, o \, X \, o \, X) - \frac{t^4}{4}(X \, o \, X \, o \, X \, o \, X) + \cdots,$$

and, in particular

$$o\text{-log}[(1 + tX_1) \, o \, (1 + tX_2)] = o\text{-log}\,(1 + tX_1) + o\text{-log}\,(1 + tX_2).$$

Now $M^*(t)$ involves brackets with one asterisk, and $M^{**}(t)$ involves brackets with two. Hence

$$M^*(t) \, o \, M^{**}(t) = M^*(t)M^{**}(t) = M_{\text{aver}}(t)$$

and, taking o-logarithms on both sides

$$o\text{-log}\, M^*(t) + o\text{-log}\, M^{**}(t) = o\text{-log}\, M_{\text{aver}}(t).$$

If we write

$$\psi(t) = (1)t + (2)\frac{t^2}{2!} + (3)\frac{t^3}{3!} + \cdots = o\text{-log}\, M(t)$$

and define $\psi_{\text{aver}}(t)$, $\psi^*(t)$ and $\psi^{**}(t)$ similarly, we have

$$\psi_{\text{aver}}(t) = \psi^*(t) + \psi^{**}(t)$$

and, comparing coefficients

$$((j)) = (j)^* + (j)^{**}$$

where $((j))$ is the same function of the $\langle\langle a\rangle\rangle$ as $(j)^*$ is of the $\langle a\rangle^*$ and $(j)^{**}$ is of the $\langle a\rangle^{**}$. Thus we have defined $(j) = k_j$ so as to have the right property. We have only to calculate the relations explicitly.

To do this, we have only to write out

$$\psi(t) = o\text{-}\log M(t)$$

remembering to use o-multiplication on the right. We find

$$\begin{aligned} tk_1 + \frac{t^2}{2!}k_2 + \frac{t^3}{3!}k_3 + \cdots &= t\langle 1\rangle + \frac{t^2}{2!}\langle 2\rangle + \frac{t^3}{3!}\langle 3\rangle + \cdots \\ &\quad -\frac{1}{2}\left(t\langle 1\rangle + \frac{t^2}{2!}\langle 2\rangle + \cdots\right) o \left(t\langle 1\rangle + \frac{t^2}{2!}\langle 2\rangle + \cdots\right) \\ &\quad + \tfrac{1}{3}(t\langle 1\rangle + \cdots)\, o\, (t\langle 1\rangle + \cdots)\, o\, (t\langle 1\rangle + \cdots) \\ &\quad \cdots \\ &= t\langle 1\rangle + \frac{t^2}{2!}\langle 2\rangle + \frac{t^3}{3!}\langle 3\rangle + \cdots \\ &\quad -\tfrac{1}{2}(t^2\langle 11\rangle + t^3\langle 12\rangle + \cdots) \\ &\quad + \tfrac{1}{3}(t^3\langle 111\rangle + \cdots) + \cdots \\ &= t\langle 1\rangle + \frac{t^2}{2!}(\langle 2\rangle - \langle 11\rangle) + \frac{t^3}{3!}(\langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle) + \cdots + , \end{aligned}$$

so that $k_1 = \langle 1\rangle$, $k_2 = \langle 2\rangle - \langle 11\rangle$, $k_3 = \langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle$, $\cdots$. In case the population is infinite, the symmetric means become moment products, the k's become cumulants and the o-multiplication becomes ordinary multiplication. These formulas become the well-known relations connecting cumulants and moments.

$$\begin{aligned} k_1 &= \mu_1', \\ k_2 &= \mu_2' - \mu_1'^2, \\ k_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3, \end{aligned}$$

and the coefficients up to order 12 are given in Kendall ([4], section 3.13).

12. Commutativity. We now wish to show why o-multiplication is commutative with additive pairing. We recall (from Section 4) that

$$\text{aver } \{\langle j\rangle\} = \langle\langle j\rangle\rangle = \sum \binom{j}{k} \langle j-k\rangle^* \langle k\rangle^{**} = \sum \binom{j}{k} \langle j-k\rangle^*\, o\, \langle k\rangle^{**}$$

and proceed to find the corresponding formula for a two-part bracket. We have

$$\langle gj\rangle = \frac{1}{n(n-1)} \sum^{\neq} (x_i^* + x_{\pi(i)}^{**})^g (x_j^* + x_{\pi(j)}^{**})^j$$

$$= \frac{1}{n(n-1)} \sum\sum \binom{g}{h}\binom{j}{k} \sum^{\neq} (x_i^*)^{g-h}(x_j^*)^{j-k}(x_{\pi(i)}^{**})^h(x_{\pi(j)}^{**})^k$$

and when we average over a random pairing, we may split the x^* from the x^{**}, just as for one-part brackets, obtaining

$$\langle\langle gj\rangle\rangle = \text{aver } \{\langle gj\rangle\} = \sum\sum \binom{g}{h}\binom{j}{k} \langle g-h, j-k\rangle^* \langle hk\rangle^{**}$$

Now

$$\langle gj\rangle = \langle g\rangle \, o \, \langle j\rangle, \qquad \langle g-h, j-k\rangle^* = \langle g-h\rangle^* \, o \, \langle j-k\rangle^*,$$
$$\langle hk\rangle^{**} = \langle h\rangle^{**} \, o \, \langle k\rangle^{**},$$

and this becomes

$$\langle\langle g\rangle \, o \, \langle j\rangle\rangle = \sum\sum \binom{g}{h}\binom{j}{k} \langle g-h\rangle^* \, o \, \langle h\rangle^{**} \, o \, \langle j-k\rangle^* \, o \, \langle k\rangle^{**}$$

$$= \left[\sum \binom{g}{h} \langle g-h\rangle^* \, o \, \langle h\rangle^{**}\right] o \left[\sum \binom{j}{k} \langle j-k\rangle^* \, o \, \langle k\rangle^{**}\right]$$

$$= \langle\langle g\rangle\rangle \, o \, \langle\langle j\rangle\rangle.$$

Thus averaging over random pairing commutes with o-multiplication for two-part brackets.

An entirely analogous proof holds for brackets with more than two parts, and, since the o-multiplication was defined for brackets and extended by linearity, we have commutativity in general. In particular, we have commutativity for polykays, so that

$$((1) \, o \, (2)) = \text{aver } \{(1) \, o \, (2)\} = ((1)) \, o \, ((2))$$
$$((g) \, o \, (j)) = \text{aver } \{(g) \, o \, (j)\} = ((g)) \, o \, ((j))$$

a result we will use almost at once.

13. The multipart k**'s.** We shall now define the multipart k's by symbolic multiplication, putting

$$(12) = k_{12} = (1) \, o \, (2) = k_1 \, o \, k_2,$$

$$(abc \cdots e) = (a) \, o \, (b) \, o \, (c) \, o \cdots o \, (e),$$

this means, of course, that we may find the expressions for the multipart k's by writing out the corresponding single-part k's in terms of brackets and symbolically multiplying out. Thus

$$(22) = (2) \, o \, (2) = [\langle 2\rangle - \langle 11\rangle] \, o \, [\langle 2\rangle - \langle 11\rangle] = \langle 22\rangle - 2\langle 112\rangle + \langle 1111\rangle.$$

We notice that, for the case of additive random pairing

$$\begin{aligned}((12)) &= ((1))\,o\,((2)) = [(1)^* + (1)^{**}]\,o\,[(2)^* + (2)^{**}] \\ &= (1)^*\,o\,(2)^* + (1)^*\,o\,(2)^{**} + (1)^{**}\,o\,(2)^* + (1)^{**}\,o\,(2)^{**} \\ &= (12)^* + (1)^*(2)^{**} + (1)^{**}(2)^* + (12)^{**}\end{aligned}$$

and that the formula for $((ab))$ is entirely analogous to this. Indeed, more complex expressions of similar form hold for the more-than-two-part k's and we immediately see that all the multipart k's satisfy the previously announced pairing formulas.

To complete our transformation formulas, we need to express brackets in terms of polykays. To this end, we write out

$$M(t) = o\text{-exp}\,(\psi(t)),$$

we find

$$\begin{aligned}1 + t\langle 1\rangle + \frac{t^2}{2!}\langle 2\rangle + \frac{t^3}{3!}\langle 3\rangle + \cdots\ 1 + tk_1 + \frac{t^2}{2!}k_2 + \frac{t^3}{3!}k_3 + \cdots \\ + \frac{1}{2}\left(tk_1 + \frac{t^2}{2!}k_2 + \cdots\right) o \left(tk_1 + \frac{t^2}{2!}k_2 + \cdots\right) \\ + \tfrac{1}{6}(tk_1 + \cdots)\,o\,(tk_1 + \cdots)\,o\,(tk_1 + \cdots) \\ + \cdots \\ = 1 + tk_1 + \frac{t^2}{2!}k_2 + \frac{t^3}{3!}k_3 + \cdots \\ + \tfrac{1}{2}(t^2k_1\,o\,k_1 + t^3k_1\,o\,k_2 + \cdots) \\ + \tfrac{1}{6}(t^3k_1\,o\,k_1\,o\,k_1 + \cdots) + \cdots \\ = 1 + tk_1 + \frac{t^2}{2!}k_2 + \frac{t^3}{3!}k_3 + \cdots \\ + \tfrac{1}{2}(t^2k_{11} + t^3k_{12} + \cdots) \\ + \frac{1}{6}\left(t^3k_{111} + \cdots\right) + \cdots \\ = 1 + tk_1 + \frac{t^2}{2}(k_2 + k_{11}) + \frac{t^3}{3!}(k_3 + 3k_{12} + k_{111}) + \cdots\end{aligned}$$

and comparing coefficients,

$$\langle 1\rangle = k_1, \qquad \langle 2\rangle = k_2 + k_{11}, \qquad \langle 3\rangle = k_3 + 3k_{12} + k_{111}\ \cdots.$$

For an infinite population, these reduce the familiar formulas expressing moments in terms of cumulants, namely

$$\mu_1 = K_1, \qquad \mu_2 = K_2 + K_{11}, \qquad \mu_3 = K_3 + 3K_{12} + K_{111}, \cdots$$

and again these can be found up to order 12 in Kendall ([4], section 3.13). This time, however, the nature of the exponential function makes it easy to write down the coefficient of

$$k_{\alpha\alpha\cdots\alpha\beta\cdots\beta\cdots\delta} \quad \text{in} \quad \langle \alpha a + \beta b + \cdots \delta\, d \rangle.$$

It is

$$\frac{(\alpha a + \beta b + \cdots + \delta d)!}{(\alpha!)^a(\beta!)^b \cdots (\delta!)^d} \frac{1}{a!b! \cdots d!}$$

For example, the coefficient of k_{12} in $\langle 3 \rangle$ is

$$\frac{3!}{1!2!1!1!} = 3.$$

Thus individual coefficients are easily checked.

REFERENCES

[1] S. H. Abdel-Aty, "Tables of generalized k-statistics." *Biometrika* Vol. 41 (1954 pp. 253–260.

[2] F. N. David and M. G. Kendall, "Tables of Symmetric functions—Part I," *Biometrika* Vol. 36 (1949) pp. 431–439.

[3] Paul L. Dressel, "Statistical seminvarients and their estimates with particular emphasis on their relation to algebraic invariants," *Annals of Mathematical Statistics*. Vol. 11 (1940) pp. 33–57.

[4] Maurice G. Kendall, *The Advanced Theory of Statistics*. Volume 1 (1943) London. Chas. Griffin and Co.

[5] M. G. Kendall, "Moment-statistics in samples from a finite population" *Biometrika* Vol. 39 (1952) pp. 14–16.

[6] John W. Tukey, "Some sampling simplified," *Journal of the American Statistical Association*. Vol. 45 (1950) 501–519.

[7] John W. Tukey, "Variances of variance components: I. Balanced designs." To appear in the *Annals of Mathematical Statistics*.

[8] John W. Tukey, "Variances of variance components: II. Unbalanced single classifications." To appear in the *Annals of Mathematical Statistics*.

[9] John W. Tukey, "Variance components: III. The third moment in a balanced single classification." To appear in the *Annals of Mathematical Statistics*.

[10] John Wishart, "Moment coefficients of the k-statistics in samples from a finite population." *Biometrika* Vol. 39, (1952) pp. 1–13.

[11] John Wishart, "The combinatorial development of the cumulants of the k-statistics," *Trabajos de Estadistica* Vol. 3 (1952) pp. 13–26.

SYMMETRIC FUNCTIONS OF A TWO-WAY ARRAY[1]

By Robert Hooke[2]

Princeton University

1. Summary. A family of polynomials in the elements of a two-way array, or matrix, is introduced. This family is an extension, from sets to matrices, of the family of symmetric polynomials k_1, k_2, k_{11}, k_3, k_{12}, etc., defined by Tukey [6], christened "polykays" in [7], and which are a generalization of the family k_1, k_2, etc., defined by R. A. Fisher [1]. The polynomials of the present paper, called "bipolykays," are symmetric functions in the sense that they are invariant under permutation of rows and/or columns of the matrix. This paper defines the bipolykays, shows that they are inherited on the average, develops the formulas for use in random pairing, and provides tables for conversion and for multiplication. A description of applications (see [2], [3], and [4]) will be postponed until a later paper. These applications include (a) finding expressions for sampling moments of functions of the elements of a matrix which is a "bisample" from a larger matrix, (b) finding expressions for sampling moments of functions (such as estimates of variance components) associated with the analysis of variance of a two-way table with systematic interactions, and (c) finding unbiased estimators for the variances and covariances of estimated variance components in a two-way table without interactions.

2. Introduction. Let $x_I(I = 1, 2, \cdots, N)$ be any population of N numbers, and let $x_i(i = 1, 2, \cdots, n)$ represent elements of a sample of size n from this population. Let $f(n; x_1, \cdots, x_n)$ be a polynomial which is symmetric in the x_i and has coefficients which are functions of n. Such a function extends obviously to a polynomial $f(N; x_1, \cdots, x_N)$, the corresponding symmetric polynomial in the x_I, with the coefficients changed only by replacing n by N. Writing "ave" for the operation of averaging over all $\binom{N}{n}$ distinct samples of size n from the population, we say that $f(n; x_1, \cdots, x_n)$ is "inherited on the average" [6] if

$$\text{ave} f(n; x_1, \cdots, x_n) = f(N; x_1, \cdots, x_N). \tag{1}$$

The functions k_1, k_2, k_{11}, etc., defined in [6] and now called polykays, are symmetric polynomials that are inherited on the average. Any symmetric polynomial can be expressed as a linear combination of polykays, so that the average value (or expected value, if the $\binom{N}{n}$ distinct samples are assigned equal probabilities) of the polynomial can be found simply by replacing each polykay in

Received July 28, 1954.

[1] Prepared in connection with research sponsored by the Office of Naval Research.

[2] Now with the Westinghouse Research Laboratories, East Pittsburgh, Pa.

this linear combination by the corresponding population polykay, i.e., by applying (1) to each term.

In order to use polykays in connection with a linear model, say

$$y_{ij} = m_i + e_{ij},$$

one needs to find the polykays of the y's in terms of those of the m's and e's. The rules for doing this are called "pairing formulas" (Section 3), and an important advantage of polykays over most other symmetric polynomial functions inherited on the average is the simplicity of their pairing formulas.

In this paper we shall consider a matrix population of numbers $x_{IJ}(I = 1, 2, \cdots, R; J = 1, 2, \cdots, C)$ from which a bisample (sample matrix) is selected by taking a sample of r of the R rows and another sample of c of the C columns and forming the matrix whose elements are at the intersections of these selected rows and columns. Symmetric polynomial functions of such matrices (i.e., polynomial functions of the elements of a matrix which are invariant under permutation of rows and/or columns) will be considered. It will be shown that any such function can be expressed as a linear combination of bipolykays, which will be defined as a special family of functions that are inherited on the average and have simple pairing formulas. These properties make the bipolykays useful in the determination of moments of moments, for example, associated with a two-way classification.

The author wishes to express here his indebtedness to Prof. J. W. Tukey for several helpful suggestions, and to Dr. Frederic Lord for posing the original matrix sampling problem (see [2]) which started this investigation.

3. Polykays. Polykays are defined by examples in [6], and a general definition may be found in [7], [8], or [9]. Since a different, though equivalent, definition appears to be more suited to the extension to bipolykays, this section will be devoted to a general definition of polykays and to the derivation of those properties which will be required in this paper. We begin with some notation and terminology of [6] which will be used throughout.

The symbol $\sum^{\neq}$ will mean a sum over all subscripts that follow, but such that subscripts represented by different letters must remain unequal throughout the summation. For example,

$$\sum_{i,j=1}^{2}{}^{\neq} x_i x_j = x_1 x_2 + x_2 x_1.$$

A *symmetric mean* is a polynomial

$$\frac{1}{M} \sum{}^{\neq} x_i^a x_j^b \cdots x_m^d,$$

where the subscripts are summed from 1 to n (for samples) or from 1 to N (for populations), the exponents are positive integers, and M is the number of terms in the summation. When the sample (or population) size is given, the symmetric

mean is specified by the exponents, and so is abbreviated by writing the exponents within brackets, as in

$$\langle a\, b\, d\rangle = \frac{1}{n(n-1)(n-2)} \sum{}^{\neq} x_i^a x_j^b x_k^d .$$

When a symmetric mean is defined over a population, this fact is indicated by a prime, as in

$$\langle a\, b\, d\rangle' = \frac{1}{N(N-1)(N-2)} \sum{}^{\neq} x_I^a x_J^b x_K^d .$$

It is obvious that any symmetric polynomial function can be expressed as a linear combination of symmetric means. Since symmetric means are inherited on the average [6], they are sufficient for the problem of finding expressions for sampling moments of moments of a single sample. However, in dealing with an additive model, one works with numbers which are sums of numbers sampled from different populations. To provide for this case, Tukey uses the notion of "random pairing": this means taking two samples, $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$, the order within each having been independently randomized, and adding the two to obtain a new sample $(z_1, \cdots, z_n)$, where $z_i = x_i + y_i$. For symmetric functions of the z's one wants the average value (where the average is taken with respect both to sampling and to randomization of order within samples) expressed, by means of a "pairing formula", in terms of symmetric functions of the two original populations. Using "ave" as before, together with "aver", meaning "average over randomization", and using one and two primes, respectively, for the populations of x's and y's, we have the following example of a pairing formula as applied to the symmetric mean $\langle 12\rangle$ taken over the z's:

$$\begin{aligned} \text{ave aver } \langle 12\rangle &= \langle 12\rangle' + \langle 1\rangle' \langle 2\rangle'' + \langle 2\rangle' \langle 1\rangle'' \\ &\quad + 2\langle 1\rangle' \langle 11\rangle'' + 2\langle 11\rangle' \langle 1\rangle'' + \langle 12\rangle''. \end{aligned}$$

The polykays are linear combinations of symmetric means chosen, among other reasons, because of their having simple pairing formulas. Those of degree 3 or less are defined as follows:

$$\begin{aligned} &k_1 = \langle 1\rangle, && k_{111} = \langle 111\rangle, \\ (2)\qquad &k_{11} = \langle 11\rangle, && k_{12} = \langle 12\rangle - \langle 111\rangle, \\ &k_2 = \langle 2\rangle - \langle 11\rangle, && k_3 = \langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle. \end{aligned}$$

The pairing formula for k_{12} becomes, for example,

$$\text{ave aver } k_{12} = k_{12}' + k_1' k_2'' + k_2' k_1'' + k_{12}'' .$$

The remainder of this section consists of a general definition of the polykays and a derivation of the pairing formulas for symmetric means and polykays. A new notation for symmetric means (to be extended later to polykays) will first

be introduced. Henceforth the notation used in (2), above, will be referred to as the *primary* notation, and that used in (3), below, as the *secondary* notation.

DEFINITION. The entries $a, b, \cdots, d$ of a symmetric mean $\langle ab \cdots d \rangle$ of degree m form a partition of the integer m. It will be convenient to represent such a partition in terms of m distinct symbols, so that the secondary notation for $\langle ab \cdots d \rangle$ will be

$$(3) \qquad \langle q_1 q_2 \cdots q_a, r_1 r_2 \cdots r_b, \cdots, s_1 s_2 \cdots s_d \rangle,$$

where commas are used to separate the *parts* of the partition, and the *lengths* of the parts are the positive integers $a, b, \cdots, d$, whose sum is m. Any use of the word partition below will refer to an expression such as that enclosed in $\langle\ \rangle$'s in (3). Two partitions are *equivalent* (not distinct) if they are identical, except possibly for the order of parts and the order of symbols within a part. Greek letters will be used to represent arbitrary partitions. A partition β is a *subpartition* of a partition α if α can be made equivalent to β merely by the insertion of one or more commas. A *dichotomy* of a partition α is an ordered set $\{\alpha_1, \alpha_2\}$ of two partitions, α_1 and α_2, such that α_1 consists of some of the symbols comprising α, and α_2 of the remaining ones, and such that any two symbols which both occur in α_1 or both in α_2 belong there to the same part if and only if they belonged to the same part of α. The null partition will be denoted by ϕ, so that $\{\phi, \alpha\}$ and $\{\alpha, \phi\}$ are dichotomies of α. A *simple* dichotomy of α into $\{\alpha_1, \alpha_2\}$ has the property that each part of α belongs entirely to α_1 or to α_2. The *join* of partitions α_1 and α_2, having no symbols in common, is that partition α such that $\{\alpha_1, \alpha_2\}$ and $\{\alpha_2, \alpha_1\}$ are simple dichotomies of α. An expression such as $\langle \alpha \rangle$, with brackets enclosing a Greek letter, will denote a symmetric mean, not with just one entry, but with entries which are the lengths of the parts of the partition α. The symmetric mean $\langle \phi \rangle$ is defined to be 1.

THEOREM 1. *The pairing formula for a symmetric mean* $\langle \alpha \rangle$ *is*

$$\text{ave aver } \langle \alpha \rangle = \sum \langle \beta \rangle' \langle \gamma \rangle'',$$

where the summation extends over all distinct dichotomies $\{\beta, \gamma\}$ *of* α.

PROOF. We recall that $\langle \alpha \rangle$ is a symmetric mean for a sample of numbers of the form $x_i + y_i$. Hence if $\langle \alpha \rangle$ is of the form (3), we have

$$\langle \alpha \rangle = \frac{1}{M} \sum\nolimits^{\neq} [(x_i + y_i) \cdots (x_i + y_i)][(x_j + y_j) \cdots (x_j + y_j)] \cdots [(x_k + y_k) \cdots (x_k + y_k)],$$

where there are a, b, and d equal factors within the first, second, and last pair of square brackets, respectively. For a fixed choice of $i, j, \cdots, k$, the product following the $\sum^{\neq}$ symbol expands to a sum of $2^{a+b+\cdots+d}$ terms, each of the form

$$(4) \qquad X_{i,j,\cdots,k} Y_{i,j,\cdots,k},$$

where

$$X_{i,j,\cdots,k} = x_i^A x_j^B \cdots x_k^D,$$

and

$$Y_{i,j,\cdots,k} = y_i^{a-A} y_j^{b-B} \cdots y_k^{d-D}.$$

Each term of the form (4) must be summed over the allowable sets of values of $i, j, \cdots, k$, averaged over randomization, and divided by M; aver $\langle\alpha\rangle$ is the sum of these individual results, one for each split of $a, b, \cdots, d$ into $A, B, \cdots, D$ and $a - A, b - B, \cdots, d - D$. From the independence of the two randomizations, we have

(5) $$\frac{1}{M} \text{aver} \sum^{\neq} X_{i,j,\cdots,k} Y_{i,j,\cdots,k} = \frac{1}{M} \sum^{\neq} \text{aver } X_{i,j,\cdots,k} \text{ aver } Y_{i,j,\cdots,k}.$$

But aver $X_{i,j,\cdots,k}$ is simply

$$\langle q_1 \cdots q_A, r_1 \cdots r_B, \cdots, s_1 \cdots s_C\rangle^* = \langle\beta\rangle^*,$$

where $\langle\beta\rangle$ is a symmetric mean of the type mentioned in the statement of the theorem, the asterisk indicating that it at present refers only to the sample of x's in question. Similarly, aver $Y_{i,j,\cdots,k}$ is $\langle\gamma\rangle^{**}$, γ and β being related as in the statement of the theorem. Hence

$$\begin{aligned} \text{aver } \langle\alpha\rangle &= \frac{1}{M} \sum \sum^{\neq} \text{aver } X_{i,j,\cdots,k} \text{ aver } Y_{i,j,\cdots,k} \\ &= \frac{1}{M} \sum \sum^{\neq} \langle\beta\rangle^* \langle\gamma\rangle^{**}. \end{aligned}$$

The M terms in the $\sum^{\neq}$ summation being equivalent, this reduces to

$$\text{aver } \langle\alpha\rangle = \sum \langle\beta\rangle^* \langle\gamma\rangle^{**},$$

this last summation being as defined in the statement of the theorem. The final step is to average over samples. Since the samples are chosen independently from different populations, and since the symmetric means are inherited on the average, we have the theorem, namely,

$$\text{ave aver } \langle\alpha\rangle = \sum \langle\beta\rangle' \langle\gamma\rangle''.$$

Definition. For partitions of a fixed number, m, of symbols, we say that

$$\text{rank } \alpha < \text{rank } \beta$$

if (a) the number of parts in α exceeds the number of parts in γ, or if (b) α and β have the same number of parts; but when the parts are arranged in order of increasing length, the first $i - 1$ parts of α are equal in length to their corresponding parts in β, while the ith part of α is shorter than the ith part of β.

Our definition of polykays will be in terms of the secondary notation, a polykay being represented by (α) and distinguishable from a symmetric mean (in this notation) only by the use of parentheses in place of $\langle\ \rangle$'s.

Definition. The polykays of degree m are defined by the equations

(6) $$\langle\alpha\rangle = (\alpha) + \sum(\beta_\alpha),$$

where there is one equation for each symmetric mean $\langle\alpha\rangle$ of degree m, and where the summation is over all distinct subpartitions β_α of α. [If there are $S(m)$ symmetric means of degree m, the $S(m)$ equations (6) of course define the $S(m)$ polykays that occur on the right if and only if the determinant of the coefficients of the distinct polykays does not vanish. (Two polykays, or two symmetric means, are equivalent, or not distinct, if the partitions representing them can be made equivalent by renaming the symbols.) Since, in each equation of (6), the rank of (α) is greater than that of any of the (β_α), then when those (β_α) which may be equivalent are collected and results are ordered by descending rank, the determinant has ones down the main diagonal and zeros below, so that its value is 1.]

Since any symmetric polynomial function can be expressed as a linear combination of symmetric means, it follows from the definition just given that it can also be expressed as a linear combination of polykays.

Example ($m = 3$). The symmetric means are $\langle 111\rangle$, $\langle 12\rangle$, and $\langle 3\rangle$, expressed in the primary notation, or $\langle p, q, s\rangle$, $\langle p, q\, s\rangle$, and $\langle p\, q\, s\rangle$ in the secondary notation, in order of ascending rank. The polykays are then defined by the equations

$$\begin{aligned}\langle p, q, s\rangle &= (p, q, s),\\ \langle p, q\, s\rangle &= (p, q\, s) + (p, q, s),\\ \langle p\, q\, s\rangle &= (p\, q\, s) + (p, q\, s) + (q, p\, s) + (s, p\, q) + (p, q, s).\end{aligned}$$

These may be solved to give

$$\begin{aligned}(p, q, s) &= \langle p, q, s\rangle,\\ (p, q\, s) &= \langle p, q\, s\rangle - \langle p, q, s\rangle,\\ (p\, q\, s) &= \langle p\, q\, s\rangle - \langle p, q\, s\rangle - \langle q, p\, s\rangle - \langle s, p\, q\rangle + 2\langle p, q, s\rangle,\end{aligned}$$

or, in the primary notation,

$$\begin{aligned}k_{111} &= \langle 111\rangle,\\ k_{12} &= \langle 12\rangle - \langle 111\rangle,\\ k_3 &= \langle 3\rangle - 3\langle 12\rangle + 2\langle 111\rangle.\end{aligned}$$

Theorem 2. *The pairing formula for a polykay* (α) *is*

$$\text{ave aver } (\alpha) = \sum (\beta)'\, (\gamma)'',$$

where the summation extends now (in contrast with the similarly written summation of Theorem 1) *only over the distinct* simple *dichotomies* $\{\beta, \gamma\}$ *of the partition* α.

Proof. We obtain the result by induction on rank for a fixed degree m. For the lowest rank, i.e., for $\langle 11 \cdots 1\rangle$ and $k_{11\cdots 1}$ (in the primary notation), the symmetric mean and polykay are identical; since all dichotomies in this case are simple, Theorem 2 holds for this rank by virtue of Theorem 1. For other ranks, we observe in equation (6) that

$$\text{ave aver } \langle\alpha\rangle = \text{ave aver } (\alpha) + \sum \text{ave aver } (\beta_\alpha) \tag{7}$$

and recall that the rank of each of the β_α is less than that of α. The induction assumption then will be that the theorem has been proved for the (β_α). Applying the theorem to any particular ave aver (β_α) gives us the sum of

$$(\gamma)'(\delta)''$$

over all distinct simple dichotomies $\{\gamma, \delta\}$ of β_α. None of these dichotomies can arise from any of the other β_α, since any simple dichotomy determines the partition from which it comes. Hence, since the various β_α are all the distinct subpartitions of α, it follows that

$$\sum \text{ave aver } (\beta_\alpha) = \sum (\gamma)'(\delta)'',$$

where the last sum extends over all $\sum (\gamma)'(\delta)''$ such that $\{\gamma, \delta\}$ is a simple dichotomy of some subpartition of α. This is the same as saying that the sum extends over all $(\gamma)'(\delta)''$ such that the join of γ and δ is a subpartition of α.

Going to the left side of (7), we have, from Theorem 1,

$$\text{ave aver } \langle\alpha\rangle = \sum \langle\lambda\rangle'\langle\mu\rangle'',$$

where the sum extends over all distinct dichotomies $\{\lambda, \mu\}$ of α. Each $\langle\lambda\rangle'$ and $\langle\mu\rangle''$ can be expressed in terms of polykays by equations (6). Since two λ's arising from distinct dichotomies of α cannot contain the same symbols, $\sum \langle\lambda\rangle'\langle\mu\rangle''$ must be equal to

$$\sum (\xi)'(\eta)''$$

where this sum extends over all terms where ξ and η are subpartitions of some λ and μ (or $\xi = \lambda$ or $\eta = \mu$ or both), respectively, $\{\lambda, \mu\}$ being a dichotomy of α. This is evidently the same as the sum of all $(\xi)'(\eta)''$ such that the join of ξ and η is α or a subpartition of α.

The first and third of the three terms of (6) have now been specified, and ave aver (α) is equal to their difference, which is the sum of all $(\xi)'(\eta)''$ such that the join of ξ and η is α.

Example. Consider the polykay k_{12}, or $(p,\ q\ s)$. The simple dichotomies of $p,\ q\ s$ are

$$\begin{array}{rcl} p,\ q\ s & \text{and} & \phi, \\ p & \text{and} & q\ s, \\ q\ s & \text{and} & p, \\ \phi & \text{and} & p,\ q\ s, \end{array}$$

so that

$$\begin{aligned} \text{ave aver } k_{12} &= \text{ave aver } (p,\ q\ s) \\ &= (p,\ q\ s)'(\phi)'' + (p)'(q\ s)'' + (q\ s)'(p)'' + (\phi)'(p,\ q\ s)'' \\ &= k'_{12} + k'_1k''_2 + k'_2k''_1 + k''_{12}, \end{aligned}$$

(ϕ) being 1.

4. Bisamples and generalized symmetric means. We turn now to the problem of the present paper. We suppose a population matrix

$$\|x_{IJ}\|, \qquad I = 1, 2, \cdots, R; \; J = 1, 2, \cdots, C$$

from which a bisample

$$\|x_{ij}\|, \qquad i = 1, 2, \cdots, r; \; j = 1, 2, \cdots, c$$

is selected as described in Section 2. Any polynomial symmetric in the x_{ij} (in the sense defined in section 2) is a linear combination of sums of the type

$$\sum\nolimits^{\neq} x_{pq}^{a_{pq}} \cdots x_{st}^{a_{st}},$$

where the symbol $\sum^{\neq}$, for two-way arrays, will mean summation over all subsequent subscripts, with the restriction that row subscripts represented by different letters must remain different throughout the summation, and the same for column subscripts.

We define *generalized symmetric means* to be averages of monomial functions over a matrix; i.e., a g.s.m. is a polynomial

$$\frac{1}{M}\Big(\sum_{p,q,\cdots,s,t}^{\neq} x_{pq}^{a_{pq}} \cdots x_{st}^{a_{st}} \Big), \tag{8}$$

where M is the number of terms in the summation. A g.s.m. is specified by the exponents, together with information which tells which ones correspond to elements that lie in the same row, and which ones correspond to elements that lie in the same column. A convenient notation for g.s.m.'s is thus provided by placing the exponents in a matrix within brackets in such a way that exponents which affect elements in the same row of the matrix $\|x_{ij}\|$ are entered in the same row, and similarly for columns. Thus,[3]

$$\begin{bmatrix} a & b & 0 \\ 0 & 0 & d \end{bmatrix} = \frac{1}{rc(r-1)(c-1)(c-2)} \sum\nolimits^{\neq} x_{ij}^{a} x_{ik}^{b} x_{mn}^{d}.$$

Ordinarily the zeros will be replaced by dashes. Dashes will also be used to extend every matrix of entries to at least two rows and two columns to avoid confusion with symmetric means and, when parentheses are later introduced, to avoid confusion with binomial coefficients. Thus,

$$\begin{bmatrix} 2 & - \\ - & - \end{bmatrix} = \frac{1}{rc} \sum\nolimits^{\neq} x_{ij}^{2}$$

$$\begin{bmatrix} 3 & - \\ 2 & - \end{bmatrix} = \frac{1}{rc(r-1)} \sum\nolimits^{\neq} x_{ij}^{3} x_{kj}^{2}.$$

Evidently two g.s.m.'s are identical if the matrix of entries of one can be obtained from that of the other by permuting rows and/or columns. The distinct g.s.m.'s of degrees 1 and 2 are as follows:

[3] Square brackets are used, for convenience in printing, in place of ⟨ ⟩'s for g.s.m.'s having more than one row.

$$\text{Degree 1: } \begin{bmatrix} 1 & - \\ - & - \end{bmatrix}.$$

$$\text{Degree 2: } \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix}, \begin{bmatrix} 1 & - \\ 1 & - \end{bmatrix}, \begin{bmatrix} 2 & - \\ - & - \end{bmatrix}.$$

The idea of random pairing for bisamples is a straightforward extension of that described in Section 3 for samples: Given two $r \times c$ matrices $\|x_{ij}\|$ and $\|y_{ij}\|$, the order of rows and of columns is randomized in each, and a new $r \times c$ matrix $\|z_{ij}\|$ is formed by matrix addition of the results.

The general term

$$x_{pq}^{a_{pq}} \cdots x_{st}^{a_{st}}$$

(which specifies, as in (8), a g.s.m. of degree m) contains m factors, a_{pq} of which are equal to x_{pq}, etc. To each of these factors we assign a different symbol, and the resulting set of symbols may be partitioned in two ways—once by rows, and once by columns. The secondary notation for the g.s.m. will then be an ordered pair

$$\langle \alpha / \beta \rangle$$

of partitions α and β, each on the same set of symbols. Each part of α will consist of those symbols which correspond to factors having a particular row subscript, and the parts of β are similarly determined by column subscripts. For example,

$$\begin{bmatrix} 2 & 1 \\ - & 1 \end{bmatrix} = \frac{1}{rc(r-1)(c-1)} \sum^{\neq} x_{pq}^{2} x_{ps} x_{ts}$$

becomes, in the secondary notation,

$$\langle a\, b\, d,\, e \;/\; a\, b,\, d\, e \rangle.$$

To establish the property of inheritance on the average, let

$$\langle \alpha / \beta \rangle' = \frac{1}{M} \sum^{\neq} x_{PQ}^{a_{PQ}} \cdots x_{ST}^{a_{ST}}$$

represent any g.s.m. for an $R \times C$ population. This is the average, with equal weights, of all terms $B = x_{PQ}^{a_{PQ}} \cdots x_{ST}^{a_{ST}}$. If $\langle \alpha / \beta \rangle$ represents the same g.s.m. for an $r \times c$ bisample, one or more of the expansions of $\langle \alpha / \beta \rangle$ over various bisamples will contain any given term B. Hence ave $\langle \alpha / \beta \rangle$ is a weighted average of all terms B, and it follows from the symmetry of the set of all $r \times c$ bisamples that the weights in this average are also equal, so that ave $\langle \alpha / \beta \rangle = \langle \alpha / \beta \rangle'$.

THEOREM 3. *For any g.s.m.* $\langle \alpha / \beta \rangle$, *the pairing formula is*

$$\text{ave aver } \langle \alpha / \beta \rangle = \sum \langle \gamma / \delta \rangle' \langle \lambda / \mu \rangle'',$$

where the summation extends over all distinct dichotomies $\{\gamma, \lambda\}$ *of* α *and* $\{\delta, \mu\}$ *of* β, γ *and* δ *consisting of the same symbols.*

The proof of this theorem, being virtually identical with that of Theorem 1, will be omitted.

5. Definition of the bipolykays. In order to make the general definition of bipolykays, we define a "dot-multiplication" for symmetric means as follows:

$$\begin{aligned} \langle\alpha\rangle\cdot\langle\beta\rangle &= \langle\alpha/\beta\rangle \quad \text{if } \alpha \text{ and } \beta \text{ consist of the same symbols} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

This noncommutative multiplication can be extended by distributivity to provide dot-products of linear combinations of symmetric means.

Definition. The *bipolykay* (α/β), where α and β are partitions of the same set of symbols, is

$$(\alpha/\beta) = (\alpha)\cdot(\beta),$$

it being understood that (α) and (β) are expressed as sums of symmetric means (as in the example just before Theorem 2, Section 3) before the dot-product is taken.

Example. Consider the bipolykay $\begin{pmatrix} 1 & 1 \\ - & 1 \end{pmatrix}$. (The primary notation for a bipolykay (α/β) is the same as that for the g.s.m. $\langle\alpha/\beta\rangle$, with $\langle\ \rangle$'s replaced by parentheses.) This becomes, in the secondary notation,

$$\begin{aligned} \begin{pmatrix} 1 & 1 \\ - & 1 \end{pmatrix} &= (p\ q,\ s\ /\ p,\ q\ s) \\ &= (p\ q,\ s)\cdot(p,\ q\ s) \qquad \text{by the definition above} \\ &= [\langle s,\ p\ q\rangle - \langle s,\ p,\ q\rangle]\cdot[\langle p,\ q\ s\rangle - \langle p,\ q,\ s\rangle] \\ &\qquad \text{by the example preceding Theorem 2} \\ &= \langle s,\ p\ q\ /\ p,\ q\ s\rangle - \langle s,\ p,\ q\ /\ p,\ q\ s\rangle - \langle s,\ p\ q\ /\ p,\ q,\ s\rangle + \langle s,\ p,\ q\ /\ p,\ q,\ s\rangle \\ &= \begin{bmatrix} 1 & 1 \\ - & 1 \end{bmatrix} - \begin{bmatrix} 1 & - \\ 1 & - \\ - & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 & - \\ - & - & 1 \end{bmatrix} + \begin{bmatrix} 1 & - & - \\ - & 1 & - \\ - & - & 1 \end{bmatrix}. \end{aligned}$$

Since bipolykays are linear combinations (with constant coefficients) of the g.s.m.'s, the bipolykays must also be inherited on the average. By means of the device of ranking (as was done for polykays in Section 3), one can show that the g.s.m.'s can in turn be expressed as linear combinations of bipolykays. (This is done explicitly through degree 4 in Section 8.) Hence any polynomial symmetric function of elements of a bisample can be expressed as a linear combination of bipolykays.

6. Pairing formulas for bipolykays. The statement of pairing formulas for bipolykays requires the following terminology:

Definition. The bipolykay (α/β) is said to be *decomposable* if there exist simple

dichotomies $\{\alpha_1, \alpha_2\}$ and $\{\beta_1, \beta_2\}$ of α and β, respectively, such that α_1 and β_1 consist of the same symbols, and neither α_1 nor α_2 is null. In this case, (α/β) may be written as a product

$$(\alpha/\beta) = (\alpha_1, \beta_1) \times (\alpha_2, \beta_2),$$

where the commutative operation denoted by $\times$ is defined by this equation, and (α_1, β_1) and (α_2, β_2) are called *components* of (α/β). If any component is similarly decomposable, the original bipolykay can be written as the $\times$-product of at least three components, and clearly any decomposable bipolykay can finally be written as the $\times$-product of indecomposable components where the set of indecomposable components is unique except for order.

THEOREM 4. *If a bipolykay* (α/β) *is indecomposable, its pairing formula is simply*

$$\text{ave aver } (\alpha/\beta) = (\alpha/\beta)' + (\alpha/\beta)''.$$

If (α, β) *is decomposable and is the* $\times$*-product of indecomposable components* (α_i/β_i), $i = 1, 2, \cdots, d$, *the pairing formula is*

$$\text{ave aver } (\alpha/\beta) = (\alpha/\beta)' + (\alpha/\beta)'' + \sum (\gamma/\delta)'(\lambda/\mu)'',$$

where the summation extends over all expressions for which (γ/δ) *is the* $\times$*-product of* $1, 2, \cdots,$ *or* $d - 1$ *of the* (α_i/β_i) *and* (λ/μ) *is the* $\times$*-product of the remaining ones.*

EXAMPLES:

$$\text{ave aver} \begin{pmatrix} 1 & 1 \\ - & 1 \end{pmatrix} = \text{ave aver } (p\ q,\ s\ /\ p,\ q\ s)$$

$$= (p\ q,\ s\ /\ p,\ q\ s)' + (p\ q,\ s\ /\ p,\ q\ s)'' \text{ since this is indecomposable}$$

$$= \begin{pmatrix} 1 & 1 \\ - & 1 \end{pmatrix}' + \begin{pmatrix} 1 & 1 \\ - & 1 \end{pmatrix}''$$

$$\text{ave aver} \begin{pmatrix} 1 & 1 & - \\ - & - & 1 \end{pmatrix} = \text{ave aver } (p\ q,\ s\ /\ p,\ q,\ s)$$

$$= \text{ave aver } [(p\ q\ /\ p,\ q) \times (s/s)]$$

$$= (p\ q,\ s\ /\ p,\ q,\ s)' + (p\ q,\ s\ /\ p,\ q,\ s)''$$

$$+ (p\ q\ /\ p,\ q)'(s/s)'' + (s/s)'(p\ q\ /\ p,\ q)''$$

$$= \begin{pmatrix} 1 & 1 & - \\ - & - & 1 \end{pmatrix}' + \begin{pmatrix} 1 & 1 & - \\ - & - & 1 \end{pmatrix}'' + \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}' \begin{pmatrix} 1 & - \\ - & - \end{pmatrix}''$$

$$+ \begin{pmatrix} 1 & - \\ - & - \end{pmatrix}' \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}''.$$

(Note: A decomposable bipolykay in primary notation can easily be recognized, as its matrix of entries can be put in the form

$$\left\| \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right\|,$$

where A, B, C, D are matrices, with all elements of B and C zero.)

The remainder of this section will be given over to the proof of Theorem 4. As before, asterisks will indicate bipolykays (or g.s.m.'s) for bisamples, and primes will denote population values; if a certain population is indicated by n primes, a bisample from that population will be indicated by n asterisks.

Definition (Extending the Dot-multiplication). In dealing with two different populations, we define

$$[\langle\alpha\rangle^*\langle\beta\rangle^{**}]\cdot[\langle\gamma\rangle^*\langle\delta\rangle^{**}] = [\langle\alpha\rangle\cdot\langle\gamma\rangle]^*[\langle\beta\rangle\cdot\langle\delta\rangle]^{**},$$

and by extension this provides a meaning for any expression which is formally written as a dot product of linear combinations of terms of the type $\langle\alpha\rangle^*\langle\beta\rangle^{**}$. Asterisks may be replaced by primes. (Note: Since we are dealing with matrices, the terms $\langle\alpha\rangle^*\langle\beta\rangle^{**}$ themselves have no meaning.)

Lemma 1. ave aver $[\langle\alpha\rangle\cdot\langle\beta\rangle]$ = [ave aver $\langle\alpha\rangle$]$\cdot$[ave aver $\langle\beta\rangle$], and ave aver $[(\alpha)\cdot(\beta)]$ = [ave aver (α)]$\cdot$[ave aver (β)]. (Here $\langle\alpha\rangle\cdot\langle\beta\rangle = \langle\alpha/\beta\rangle$ is a g.s.m. for a sample formed by random pairing of two bisamples. The expressions ave aver $\langle\alpha\rangle$ and ave aver $\langle\beta\rangle$ are formal expressions of Theorem 1, their dot product having meaning only from the definition just above. Similarly for the polykays, which must be expressed as sums of g.s.m.'s before the above definition gives them meaning.)

Proof. If α and β do not consist of the same symbols, the result is trivial. If they do, then $\langle\alpha\rangle\cdot\langle\beta\rangle = \langle\alpha/\beta\rangle$, and so

$$\text{ave aver } [\langle\alpha\rangle\cdot\langle\beta\rangle] = \sum \langle\gamma/\delta\rangle'\langle\lambda/\mu\rangle'',$$

where γ, δ, λ, μ are as described in Theorem 3. Clearly, from Theorem 1, [ave aver $\langle\alpha\rangle$]$\cdot$[ave aver $\langle\beta\rangle$] gives the same sum, so we may now go to the second part of this lemma. In the case of polykays, we have, by their definition,

$$(\alpha) = \textstyle\sum_i a_i\langle\alpha_i\rangle,$$
$$(\beta) = \textstyle\sum_j b_j\langle\beta_j\rangle,$$

where α_i is α or a subpartition of α, and the β_j bear the same relation to β. Hence

$$\text{ave aver } [(\alpha)\cdot(\beta)] = \textstyle\sum_i \sum_j a_i b_j \text{ ave aver } [\langle\alpha_i\rangle\cdot\langle\beta_j\rangle],$$

and

$$[\text{ave aver } (\alpha)]\cdot[\text{ave aver } (\beta)] = \textstyle\sum_i \sum_j a_i b_j \,[\text{ave aver } \langle\alpha_i\rangle]\cdot[\text{ave aver } \langle\beta_j\rangle].$$

By the first part of the lemma these are the same, and so the second part is proved.

LEMMA 2. $(\alpha)'(\beta)''\cdot(\gamma)'(\delta)'' = [(\alpha)\cdot(\gamma)]'[(\beta)\cdot(\delta)]''$. *(Each side of this equation has the meaning that is provided by the above conventions after each polykay has been written as a linear combination of symmetric means.)*

PROOF. As in Lemma 1, we write

$$(\alpha)' = \sum_i a_i\langle\alpha_i\rangle',$$
$$(\beta)'' = \sum_j b_j\langle\beta_j\rangle'',$$
$$(\gamma)' = \sum_k c_k\langle\gamma_k\rangle',$$
$$(\delta)'' = \sum_m d_m\langle\delta_m\rangle''.$$

Then

$$\begin{aligned}(\alpha)'(\beta)''\cdot(\gamma)'(\delta)'' &= \sum_i \sum_j a_ib_j\langle\alpha_i\rangle'\langle\beta_j\rangle''\cdot\sum_k\sum_m c_k\, d_m\langle\gamma_k\rangle'\langle\delta_m\rangle'' \\ &= \sum_i \sum_j \sum_k \sum_m a_ib_jc_k\, d_m\langle\alpha_i\rangle'\langle\beta_j\rangle''\cdot\langle\gamma_k\rangle'\langle\delta_m\rangle'' \\ &\qquad\qquad\text{by definition} \\ &= \sum_i \sum_j a_ib_j[\langle\alpha_i\rangle\cdot\langle\gamma_k\rangle]' \sum_k \sum_m [\langle\beta_j\rangle\cdot\langle\delta_m\rangle]'' \\ &= [(\alpha)\cdot(\gamma)]'[(\beta)\cdot(\delta)]'',\end{aligned}$$

proving Lemma 2.

To prove Theorem 4, we write

$$\begin{aligned}\text{ave aver } (\alpha/\beta) &= \text{ave aver } [(\alpha)\cdot(\beta)] && \text{by definition} \\ &= [\text{ave aver } (\alpha)]\cdot[\text{ave aver } (\beta)] && \text{by Lemma 1} \\ &= \sum (\lambda_\alpha)'(\mu_\alpha)''\cdot \sum (\lambda_\beta)'(\mu_\beta)'' && \text{by Theorem 2,}\end{aligned}$$

where the first sum extends over all simple dichotomies $\{\lambda_\alpha, \mu_\alpha\}$ of α, and similarly for the second sum. Hence

$$\text{ave aver } (\alpha/\beta) = \sum\sum [(\lambda_\alpha)'\cdot(\lambda_\beta)'][(\mu_\alpha)''\cdot(\mu_\beta)''],$$

by Lemma 2. Now λ_α is a partition consisting of some of the parts of α, with no other changes made, since $\{\lambda_\alpha, \mu_\alpha\}$ is a simple dichotomy of α. Similarly λ_β consists of some of the parts of β. The expression $(\lambda_\alpha)'\cdot(\lambda_\beta)'$ vanishes unless λ_α and λ_β comprise exactly the same symbols. Thus the only nonvanishing terms arise when

(a) $\lambda_\alpha = \alpha$ and $\lambda_\beta = \beta$, producing the term

$$[(\lambda_\alpha)'\cdot(\lambda_\beta)'][(\phi)''\cdot(\phi)''] = (\alpha/\beta)',$$

ϕ being null; or

(b) $\lambda_\alpha = \lambda_\beta = \phi$, producing the term $(\alpha/\beta)''$; or

(c) $\lambda_\alpha = \lambda_\beta$ and $\mu_\alpha = \mu_\beta$ and none of these is null.

The last case cannot happen to an indecomposable bipolykay. If the bipolykay is decomposable, case (c) gives exactly the various terms that correspond to the splitting of the bipolykay into indecomposable components, and Theorem 4 is established.

7. Pairing formulas for certain special cases. Various special cases and degenerate cases arise in connection with pairing when applied to the analysis of variance. In order to deal with some of these, we need first a lemma about polykays and then a theorem:

LEMMA 3. *If $k_{mn\cdots p}$ is any polykay, the coefficients in its expression as a linear combination of symmetric means add to* 0 *unless $m = n = \cdots = p = 1$.*

PROOF. Consider a population, or sample, all of whose elements are equal to 1. Then clearly every symmetric mean has the value 1, and the value of $k_{mn\cdots p}$ is the sum of the coefficients in its expression as a linear combination of symmetric means. We have only to show that in this case $k_{mn\cdots p} = 0$ unless

$$m = n = \cdots = p = 1.$$

Looking at equation (6), we see that, when all parts of α are of length 1, $\langle\alpha\rangle = (\alpha)$, i.e.,

$$k_{11\cdots 1} = \langle 11 \cdots 1\rangle,$$

and, in the present case, each of these equals 1. If α has one part of length 2, and all others are of length 1, then

$$\langle 11 \cdots 12\rangle = (11 \cdots 12) + (11 \cdots 1);$$

or $1 = (11 \cdots 12) + 1$ in the present case, so that $(11 \cdots 12)$ is 0. We can now prove the theorem by induction on rank, supposing that, for a given equation of type (6),

$$\langle\alpha\rangle = (\alpha) + \sum (\beta_a),$$

all polykays of rank less than that of (α) are 0 except for $(11 \cdots 1)$. This equation then becomes

$$1 = (\alpha) + 1,$$

and $(\alpha) = 0$.

THEOREM 5. *Consider a bisample in which all elements in the same row are equal, i.e.,*

$$x_{ij} = x_i \qquad j = 1, 2, \cdots, c.$$

Over this matrix, a bipolykay (α/β) (a) *is equal to the polykay* (α), *defined over the set of x_i, if all parts of the partition β are of length* 1 *(i.e., if, in the primary notation, all entries are ones in different columns)*; or (b) *is equal to* 0 *otherwise.* Obviously an analogous statement applies to a bisample with constant columns.

PROOF. In this case it is obvious that any g.s.m. $\langle\alpha/\beta\rangle$ is equal to $\langle\alpha\rangle$, the latter being defined over the set of x_i. Now if (α/β) is any polykay, we can write

$$\begin{aligned}(\alpha/\beta) &= (\alpha)\cdot(\beta)\\ &= \textstyle\sum_i a_i\langle\alpha_i\rangle\cdot\sum_j b_j\langle\beta_j\rangle && \text{by definition}\\ &= \textstyle\sum_i\sum_j a_ib_j\langle\alpha_i/\beta_j\rangle\\ &= \textstyle\sum_i\sum_j a_ib_j\langle\alpha_i\rangle, && \text{by remark at beginning of this paragraph.}\end{aligned}$$

This last expression vanishes unless $\sum b_j \neq 0$, i.e., unless (Lemma 3) all parts of the partition β are of length 1. Hence $\sum b_j$ is 1 when it is not 0, and in this case

$$\begin{aligned}(\alpha/\beta) &= \textstyle\sum_i a_i\langle\alpha_i\rangle\\ &= (\alpha_i).\end{aligned}$$

The special cases which we now wish to consider are as follows:

CASE I (CONSTANT ROWS). Theorem 5 shows that in this case a bipolykay (α/β) is given by

$$\begin{aligned}(\alpha/\beta) &= k_{mn\cdots p} && \text{if all parts of } \beta \text{ are of length 1}\\ &= 0 && \text{otherwise,}\end{aligned}$$

$m, n, \cdots, p$ being the lengths of the parts of α.

CASE II (CONSTANT COLUMNS). Here, of course,

$$\begin{aligned}(\alpha/\beta) &= k_{mn\cdots p} && \text{if all parts of } \alpha \text{ are of length 1}\\ &= 0 && \text{otherwise.}\end{aligned}$$

CASE III (CONSTANT ROWS AND COLUMNS). Here all elements of the bisample are equal. It follows that

$$\begin{aligned}(\alpha/\beta) &= k_{11\cdots 1} && \text{if all parts of } \alpha \text{ and } \beta \text{ have length 1}\\ &= 0 && \text{otherwise.}\end{aligned}$$

If d is the common value of the elements of the bisample, and m is the degree of (α/β), then $(\alpha/\beta) = d^m$ when it does not vanish.

These cases might arise, for example, in connection with a linear model such as

$$u_{ij} = m + x_i + y_i + z_{ij},$$

where the x_i can be thought of as a bisample from a matrix with constant rows, the y_j from a matrix with constant columns, m from a matrix with all elements equal, and z_{ij} from an arbitrary matrix representing "cell effects". Using 0, 1, 2, 3, and 4 primes (asterisks) for the populations (bisamples) respectively associ-

ated with u, m, x, y, and z, we find the pairing formula for any indecomposable bipolykay (α/β), for example, to be

$$\text{ave aver } (\alpha/\beta) = (\alpha/\beta)' + (\alpha/\beta)'' + (\alpha/\beta)''' + (\alpha/\beta)'''',$$

and the cases above would tell us that some of these terms are zero and others are equivalent to polykays, depending on what bipolykay (α/β) represents.

Instead of sampling from a matrix, one may wish to consider a degenerate case in which the population consists only of rows, with no column designations; i.e., an $r \times c$ bisample is chosen by a selection of r rows followed by a selection of c elements from each of the r rows chosen. There is also the completely degenerate case, with no rows or columns, so that an $r \times c$ bisample is just an ordinary sample, randomly arranged, of rc elements from a set of numbers. This case, which would apply, for example, to the z's in the linear model above if they were regarded as independently sampled "random errors" instead of fixed interactions, is designated as Case IV:

Case IV. When pairing a bisample with a completely degenerate bisample, we have only to notice that randomization in the latter is not restricted to rows and columns, so that we have, for the completely degenerate case, these results:

(a) All g.s.m.'s with the same entries (primary notation) have equal averages for randomization, e.g.,

$$\text{aver}\begin{bmatrix} 2 & 1 \\ - & - \end{bmatrix} = \text{aver}\begin{bmatrix} 2 & - \\ 1 & - \end{bmatrix} = \text{aver}\begin{bmatrix} 2 & - \\ - & 1 \end{bmatrix}.$$

(b) All bipolykays vanish on the average except those having only diagonal elements (in secondary notation, this means those (α/β) such that α and β are equivalent partitions). This statement can be verified for degrees ≤ 4 by observing that, in the relevant conversion formulas (Section 8), the coefficients of g.s.m.'s with the same entries add to zero.

(c) Bipolykays with only diagonal entries are equal, on the average, to the corresponding polykays; e.g.,

$$\text{aver}\begin{pmatrix} 2 & - \\ - & 1 \end{pmatrix} = \text{aver } k_{21}, \qquad \text{etc.}$$

8. Conversion formulas for g.s.m.'s and bipolykays. In this and the next section, tables will be presented which make possible the use of bipolykays up through degree 4, that is, up through variances of variances. The distinct g.s.m.'s of degrees 1 and 2 were listed in Section 4. Those of degrees 3 and 4 require more space (in either notation) and so will be denoted by t's (for degree 3) and f's (for degree 4) with subscripts, even though this notation is less informative, as follows:

Degree 3: $$t_1 = \begin{bmatrix} 1 & - & - \\ - & 1 & - \\ - & - & 1 \end{bmatrix} \qquad t_6 = \begin{bmatrix} 1 & 1 \\ 1 & - \end{bmatrix}$$

$$t_2 = \begin{bmatrix} 1 & 1 & - \\ - & - & 1 \end{bmatrix} \qquad t_7 = \begin{bmatrix} 2 & - \\ - & 1 \end{bmatrix}$$

$$t_3 = \begin{bmatrix} 1 & - \\ 1 & - \\ - & 1 \end{bmatrix} \qquad t_8 = \begin{bmatrix} 2 & 1 \\ - & - \end{bmatrix}$$

$$t_4 = \begin{bmatrix} 1 & 1 & 1 \\ - & - & - \end{bmatrix} \qquad t_9 = \begin{bmatrix} 2 & - \\ 1 & - \end{bmatrix}$$

$$t_5 = \begin{bmatrix} 1 & - \\ 1 & - \\ 1 & - \end{bmatrix} \qquad t_{10} = \begin{bmatrix} 3 & - \\ - & - \end{bmatrix}$$

Degree 4:

$$f_1 = \begin{bmatrix} 1 & - & - & - \\ - & 1 & - & - \\ - & - & 1 & - \\ - & - & - & 1 \end{bmatrix} \qquad f_{12} = \begin{bmatrix} 2 & - & - \\ - & 1 & - \\ - & - & 1 \end{bmatrix} \qquad f_{23} = \begin{bmatrix} 2 & - \\ 1 & - \\ 1 & - \end{bmatrix}$$

$$f_2 = \begin{bmatrix} 1 & 1 & - & - \\ - & - & 1 & - \\ - & - & - & 1 \end{bmatrix} \qquad f_{12} = \begin{bmatrix} 1 & 1 & - \\ 1 & - & 1 \end{bmatrix} \qquad f_{24} = \begin{bmatrix} 2 & 1 \\ - & 1 \end{bmatrix}$$

$$f_3 = \begin{bmatrix} 1 & - & - \\ 1 & - & - \\ - & 1 & - \\ - & - & 1 \end{bmatrix} \qquad f_{14} = \begin{bmatrix} 1 & 1 \\ 1 & - \\ - & 1 \end{bmatrix} \qquad f_{25} = \begin{bmatrix} 2 & - \\ 1 & 1 \end{bmatrix}$$

$$f_4 = \begin{bmatrix} 1 & 1 & - & - \\ - & - & 1 & 1 \end{bmatrix} \qquad f_{15} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & - & - \end{bmatrix} \qquad f_{26} = \begin{bmatrix} 2 & 1 \\ 1 & - \end{bmatrix}$$

$$f_5 = \begin{bmatrix} 1 & - \\ 1 & - \\ - & 1 \\ - & 1 \end{bmatrix} \qquad f_{16} = \begin{bmatrix} 1 & 1 \\ 1 & - \\ 1 & - \end{bmatrix} \qquad f_{27} = \begin{bmatrix} 2 & - \\ - & 2 \end{bmatrix}$$

$$f_6 = \begin{bmatrix} 1 & 1 & 1 & - \\ - & - & - & 1 \end{bmatrix} \qquad f_{17} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad f_{28} = \begin{bmatrix} 3 & - \\ - & 1 \end{bmatrix}$$

$$f_7 = \begin{bmatrix} 1 & - \\ 1 & - \\ 1 & - \\ - & 1 \end{bmatrix} \qquad f_{18} = \begin{bmatrix} 2 & - & - \\ - & 1 & 1 \end{bmatrix} \qquad f_{29} = \begin{bmatrix} 2 & 2 \\ - & - \end{bmatrix}$$

$$f_8 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ - & - & - & - \end{bmatrix} \qquad f_{19} = \begin{bmatrix} 2 & - \\ - & 1 \\ - & 1 \end{bmatrix} \qquad f_{30} = \begin{bmatrix} 2 & - \\ 2 & - \end{bmatrix}$$

$$f_9 = \begin{bmatrix} 1 & - \\ 1 & - \\ 1 & - \\ 1 & - \end{bmatrix} \qquad f_{20} = \begin{bmatrix} 2 & 1 & - \\ - & - & 1 \end{bmatrix} \qquad f_{31} = \begin{bmatrix} 3 & 1 \\ - & - \end{bmatrix}$$

$$f_{10} = \begin{bmatrix} 1 & 1 & - \\ 1 & - & - \\ - & - & 1 \end{bmatrix} \qquad f_{21} = \begin{bmatrix} 2 & - \\ 1 & - \\ - & 1 \end{bmatrix} \qquad f_{32} = \begin{bmatrix} 3 & - \\ 1 & - \end{bmatrix}$$

$$f_{11} = \begin{bmatrix} - & 1 & 1 \\ 1 & - & - \\ 1 & - & - \end{bmatrix} \qquad f_{22} = \begin{bmatrix} 2 & 1 & 1 \\ - & - & - \end{bmatrix} \qquad f_{33} = \begin{bmatrix} 4 & - \\ - & - \end{bmatrix}$$

The following conversion formulas apply to the bipolykays of degrees 1 and 2:

Degree 1: $$(1) = \langle 1 \rangle$$

Degree 2: $$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix} = \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix} - \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}$$

$$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix} = \begin{bmatrix} 1 & - \\ 1 & - \end{bmatrix} - \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}$$

$$\begin{pmatrix} 2 & - \\ - & - \end{pmatrix} = \begin{bmatrix} 2 & - \\ - & - \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix} - \begin{bmatrix} 1 & - \\ 1 & - \end{bmatrix} + \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}.$$

The bipolykays of degree 2 have been independently developed by H. Fairfield Smith in [5].

For degrees 3 and 4 we use notation analogous to that used above for g.s.m.'s, letting T's stand for bipolykays of degree 3 and F's for bipolykays of degree 4. Thus

$$T_4 = \begin{pmatrix} 1 & 1 & 1 \\ - & - & - \end{pmatrix}$$

$$F_6 = \begin{pmatrix} 1 & 1 & 1 & - \\ - & - & - & 1 \end{pmatrix}, \qquad \text{etc.}$$

The conversion formulas for bipolykays of degrees 3 and 4 become quite long; but since they are linear, only the coefficients are of interest. These coefficients are found in Table 1 for degree 3 and in Table 2 for degree 4. The nature of the formulas makes it possible for one table to present coefficients for conversion in both directions. The coefficients for any desired expression are found by reading over (or down) to and including the diagonal of ones. For example, in Table 1,

TABLE 1

For conversion of g.s.m.'s and bipolykays of degree 3

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
T_1	1	1	1	1	1	1	1	1	1	1
T_2	−1	1		3		1	1	3	1	3
T_3	−1		1		3	1	1	1	3	3
T_4	2	−3		1				1		1
T_5	2		−3		1				1	1
T_6	1	−1	−1			1		2	2	6
T_7	1	−1	−1				1	1	1	3
T_8	−2	3	2	−1		−2	−1	1		3
T_9	−2	2	3	.	−1	−2	−1		1	3
T_{10}	4	−6	−6	2	2	6	3	−3	−3	1

$$T_6 = t_1 - t_2 - t_3 + t_6,$$
$$t_5 = T_1 + 3T_3 + T_5,$$

and similarly in Table 2.

9. Multiplication formulas for bipolykays. The usefulness of the property of inheritance on the average is pretty well limited to the case where functions having this property occur linearly. Any polynomial in bipolykays for one bisample, however, can be expressed as a linear combination of bipolykays for that bisample, given the proper multiplication formulas. We give below the multiplication formulas for bipolykays, up to and including products of degree 4.

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^2 = \begin{pmatrix}2 & -\\ - & -\end{pmatrix} + c\begin{pmatrix}1 & 1\\ - & -\end{pmatrix} + r\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix} + rc\begin{pmatrix}1 & -\\ - & 1\end{pmatrix};$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}\begin{pmatrix}1 & -\\ - & -\end{pmatrix} = 2T_7 + 2T_6 + 2rT_3 + 2cT_2 + rcT_1,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix} = 2T_8 + 2rT_6 + cT_4 + rcT_2,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix} = 2T_9 + 2cT_6 + rT_5 + rcT_3,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}\begin{pmatrix}2 & -\\ - & -\end{pmatrix} = T_{10} + rT_9 + cT_8 + rcT_7;$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}T_1 = 6F_{10} + 3F_{12} + 3rF_3 + 3cF_2 + rcF_1,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}T_2 = 2F_{13} + F_{15} + F_{18} + 2F_{20} + r(2F_{10} + F_{11}) + c(F_4 + F_6) + rcF_2,$$

TABLE 2

For conversion of g.s.m.'s and bipolykays of degree 4

F	f																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	−1	1		2		3		6		1	1	1	2	1	3	1	2	2	1	3	1	6	1	3	2	3	2	3	6	2	6	3	6
3	−1		1		2		3		6	1	1	1	1	2	1	3	2	1	2	1	3	1	6	2	3	3	2	3	2	6	3	6	6
4	1	−2		1				3					1				1	1				3			1		1		3	1	3		3
5	1		−2		1				3					1			1		1				3	1			1		1	3		3	3
6	2	−3				1		4							1					1		4		1		1		1	4		4	1	4
7	2		−3				1		4							1					1		4		1	1		1		4	1	4	4
8	−6	12		−3		−4		1														1							1		1		1
9	−6		12		−3		−4		1														1							1		1	1
10	1	−1	−1							1			2	2	2	2	4			2	2	4	4	4	4	6		6	8	8	12	12	24
11	1	−1	−1								1				1	1		1	1			1	1	1	1	2	2		2	2	3	3	6
12	1	−1	−1									1						1	1	1	1	1	1	1	1	1	2	3	2	2	3	3	6
13	−1	2	1	−1						−2			1				2					2			2				4	4	6		12
14	−1	1	2		−1					−2				1			2						2	2					4	4		6	12
15	−2	3	2			−1				−2	−1				1							2		1		2			4		6	3	12
16	−2	2	3				−1			−2	−1					1							2		1	2				4	3	6	12
17	1	−2	−2	1	1					4			−2	−2			1												2	2			6
18	−1	2	1	−1							−1	−1						1				1			1		2		2	2	3		6
19	−1	1	2		−1						−1	−1							1				1	1			2		2	2		3	6
20	−2	3	2			−1				−2		−1								1		2		1		1		3	4		6	3	12
21	−2	2	3				−1			−2		−1									1		2		1	1		3		4	3	6	12
22	6	−12	−6	3		4		−1		8	2	2	−2		−2			−1		−2		1							2		3		6
23	6	−6	−12		3		4		−1	8	2	2		−2		−2			−1		−2		1							2		3	6
24	2	−3	−4		2	1				4	1	1		−2	−1				−1	−1				1					4			3	12
25	2	−4	−3	2			1			4	1	1	−2			−1		−1			−1				1					4	3		12
26	4	−6	−6			2	2			6	2	1			−2	−2				−1	−1										3	3	12
27	1	−2	−2	1	1						2	2						−2	−2							1	1		1	1			3
28	4	−6	−6			2	2			6		3								−3	−3							1			1	1	4
29	−6	12	12	−3	−6	−4		1		−16	−4	−4	4	8	4		−2	2	4	4		−2		−4			−1		1				3
30	−6	12	12	−6	−3		−4		1	−16	−4	−4	8	4		4	−2	4	2		4		−2		−4		−1			1			3
31	−12	24	18	−6		−8	−6	2		−24	−6	−6	6		6	6		3		6	6	−3			−3	−3		−1			1		4
32	−12	18	24		−6	−6	−8		2	−24	−6	−6		6	6	6			3	6	6		−3	−3		−3		−1				1	4
33	36	−72	−72	18	18	24	24	−6	−6	96	24	24	−24	−24	−24	−24	6	−12	−12	−24	−24	12	12	12	12	12	3	4	−3	−3	−4	−4	1

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_3 = 2F_{14} + F_{16} + F_{19} + 2F_{21} + r(F_5 + F_7) + c(2F_{10} + F_{11}) + rcF_3,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_4 = 3F_{22} + 3rF_{15} + cF_8 + rcF_6,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_5 = 3F_{23} + rF_9 + 3cF_{16} + rcF_7,$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_6 = F_{17} + F_{24} + F_{25} + F_{26} + r(F_{14} + F_{16}) + c(F_{13} + F_{15}) + rcF_{10},$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_7 = F_{24} + F_{25} + F_{27} + F_{28} + r(F_{19} + F_{21}) + c(F_{18} + F_{20}) + rcF_{12},$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_8 = F_{29} + F_{31} + r(F_{24} + F_{26}) + cF_{22} + rcF_{20},$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_9 = F_{30} + F_{32} + rF_{23} + c(F_{25} + F_{26}) + rcF_{21},$$

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix} T_{10} = F_{33} + rF_{32} + cF_{31} + rcF_{28}.$$

Products of bipolykays of degree 2 are more complicated. The coefficients of the bipolykays of degree 4 in the expressions of these products are tabulated below, using the following abbreviations:

$$\begin{array}{ll} a = 2/[rc(c-1)] & g = 1/c \\ b = 2/[rc(r-1)] & h = 1/r \\ d = 2/[c(c-1)] & k = 1/(rc) \\ e = 2/[r(r-1)] & p = 1/[rc(r-1)(c-1)] \end{array}$$

10. Variances of bipolykays of degree 2. The multiplication table of the preceding section enables us to find the variances, in taking bisamples from a population matrix, of bipolykays of degree 1 or 2. For example, we have

$$\operatorname{var}\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^* = \text{ave aver}\left\{\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^*\right\}^2 - \left\{\text{ave aver}\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^*\right\}^2.$$

From Section 9 we have

$$rc\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^2 = \begin{pmatrix}2 & -\\ - & -\end{pmatrix} + c\begin{pmatrix}1 & 1\\ - & -\end{pmatrix} + r\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix} + rc\begin{pmatrix}1 & -\\ - & 1\end{pmatrix},$$

and so

$$\begin{aligned}\text{ave aver}\left\{\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^*\right\}^2 &= \text{ave aver}\left\{\frac{1}{rc}\begin{pmatrix}2 & -\\ - & -\end{pmatrix}^* + \frac{1}{r}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}^* + \frac{1}{c}\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}^* + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}^*\right\} \\ &= \frac{1}{rc}\begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \frac{1}{r}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' + \frac{1}{c}\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}',\end{aligned}$$

TABLE 3

Products of bipolykays of degree 2

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}	F_{19}	F_{20}	F_{21}	F_{24}	F_{25}	F_{26}	F_{27}	F_{28}
$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}^2$	1	$4h$	$4g$	e	d	0	0	$8k$	$4k$	$4k$	$2b$	$2a$	0	0	$2p$	$2b$	$2a$	0	0	0	0	0	$2p$	0
$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}$	0	1	0	$-e$	0	$2h$	0	$4g$	0	0	$-2b$	d	$4k$	0	$-2p$	$-2b$	0	$4k$	0	$2a$	0	0	$-2p$	0
$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}$	0	0	1	0	$-d$	0	$2g$	$4h$	0	0	e	$-2a$	0	$4k$	$-2p$	0	$-2a$	0	$4k$	0	$2b$	0	$-2p$	0
$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$	0	0	0	0	0	0	0	0	0	1	$-e$	$-d$	0	0	$2p$	0	0	$2h$	$2g$	$-2a$	$-2b$	$2k$	$2p$	$2k$
$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}$	0	0	0	0	0	0	0	0	1	0	$-e$	$-d$	$2h$	$2g$	$2p$	0	0	0	0	$-2a$	$-2b$	$4k$	$2p$	0

	F_4	F_5	F_8	F_9	F_{13}	F_{14}	F_{17}	F_{18}	F_{19}	F_{22}	F_{23}	F_{24}	F_{25}	F_{27}	F_{29}	F_{30}	F_{31}	F_{32}	F_{33}
$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^2$	$1+re$	0	h	0	$2r^2b$	0	$2r^2p$	$2rb$	0	$4k$	0	0	0	$2rp$	a	0	0	0	0
$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}^2$	0	$1+cd$	0	g	0	$2c^2a$	$2c^2p$	0	$2ca$	0	$4k$	0	0	$2cp$	0	b	0	0	0
$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$	0	0	0	0	re	0	$-2r^2p$	1	0	h	0	0	$2rb+2g$	$-2rp$	$-a$	0	$2k$	0	0
$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$	0	0	0	0	0	cd	$-2c^2p$	0	1	0	g	$2ca+2h$	0	$-2cp$	0	$-b$	0	$2k$	0
$\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}^2$	0	0	0	0	0	0	$2(r^2c^2p-1)$	0	0	0	0	0	0	$2rcp+1$	$ac+h$	$br+g$	0	0	k

and

$$\left\{\text{ave aver}\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^{*}\right\}^2 = \left\{\begin{pmatrix}1 & -\\ - & -\end{pmatrix}'\right\}^2$$

$$= \frac{1}{RC}\begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \frac{1}{R}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' + \frac{1}{C}\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}'.$$

Hence

$$\text{var}\begin{pmatrix}1 & -\\ - & -\end{pmatrix}^{*} = \left(\frac{1}{rc} - \frac{1}{RC}\right)\begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \left(\frac{1}{r} - \frac{1}{R}\right)\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' + \left(\frac{1}{c} - \frac{1}{C}\right)\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}'.$$

Proceeding in the same way for the variances of bipolykays of degree 2, we obtain the results given in Table 4, which table provides the coefficients of the expressions of the indicated variances and covariances as linear combinations of population bipolykays. In order to simplify the tabulation, the following expressions are used:

$$A = \frac{1}{r} - \frac{1}{R} \qquad S = \frac{1}{rc} - \frac{1}{RC}$$

$$B = \frac{1}{c} - \frac{1}{C} \qquad T = \frac{2}{(r-1)(c-1)} - \frac{2}{(R-1)(C-1)}$$

$$D = \frac{2}{r-1} - \frac{2}{R-1} \qquad U = \frac{2}{rc(r-1)} - \frac{2}{RC(R-1)}$$

$$E = \frac{2}{c-1} - \frac{2}{C-1} \qquad V = \frac{2}{rc(c-1)} - \frac{2}{RC(C-1)}$$

$$G = \frac{2}{r(r-1)} - \frac{2}{R(R-1)} \qquad W = \frac{2}{c(r-1)(c-1)} - \frac{2}{C(R-1)(C-1)}$$

$$H = \frac{2}{c(c-1)} - \frac{2}{C(C-1)} \qquad Y = \frac{2}{r(r-1)(c-1)} - \frac{2}{R(R-1)(C-1)}$$

$$P = \frac{2}{c(r-1)} - \frac{2}{C(R-1)} \qquad Z = \frac{2}{rc(r-1)(c-1)} - \frac{2}{RC(R-1)(C-1)}.$$

$$Q = \frac{2}{r(c-1)} - \frac{2}{R(C-1)}$$

11. Conclusion. Any symmetric function of elements of a two-way array can be expressed as a linear combination of bipolykays, using the multiplication formulas of Section 9 where necessary. If there is a linear model involved, then the average values of the bipolykays (and hence of the original symmetric function) can be found, by means of pairing formulas, in terms of polykays or bipolykays of the populations from which come the components of the linear model. A later paper will illustrate the use of these procedures in finding unbiased estimators for variance components, as well as the variances of these estimators, etc.

TABLE 4

Variances and covariances of $\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}$, $\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}$, *and* $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}$

	F_4'	F_5'	F_8'	F_9'	F_{13}'	F_{14}'	F_{17}'	F_{18}'	F_{19}'	F_{22}'	F_{23}'	F_{27}'	F_{29}'	F_{30}'	F_{32}'
var $\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}$	D	0	A	0	$4B+2P$	0	$H+W$	$2P$	0	$4S$	0	W	V	0	0
var $\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}$	0	E	0	B	0	$4A+2Q$	$G+Y$	0	$2Q$	0	$4S$	Y	0	U	0
var $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}$	0	0	0	0	0	0	$D+E+T$	0	0	0	0	T	$A+Q$	$B+P$	S

	F_{13}'	F_{14}'	F_{15}'	F_{16}'	F_{17}'	F_{22}'	F_{23}'	F_{2}'	F_{25}'	F_{26}'	F_{27}'	F_{29}'	F_{30}'	F_{31}'	F_{32}'
cov $\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}$, $\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}$	$-G$	$-H$	$2A$	$2B$	Z	0	0	$-2V$	$-2U$	$4S$	Z	0	0	0	0
cov $\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}$, $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}$	D	0	0	0	$-W-H$	A	0	0	$2B+2P$	0	$-W$	$-V$	0	$2S$	0
cov $\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}$, $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}$	0	E	0	0	$-Y-G$	0	B	$2A+2Q$	0	0	$-Y$	0	$-U$	0	$2S$

REFERENCES

[1] R. A. FISHER, "Moments and product moments of sampling distributions," *Proc. London Math. Soc.* (2), Vol. 30 (1928–29), pp. 199–238.

[2] R. HOOKE, "Sampling from a matrix, with applications to the theory of testing," Statistical Research Group, Princeton University, Memorandum Report 53, Nov., 1953.

[3] R. HOOKE, "Moments of moments in matrix sampling—an extension of polykays," Statistical Research Group, Princeton University, Memorandum Report 55, April, 1954.

[4] R. HOOKE, "The estimation of polykays in the analysis of variance," Statistical Research Group, Princeton University, Memorandum Report 56, May, 1954.

[5] H. FAIRFIELD SMITH, "Variance components, finite populations, and experimental inference," University of North Carolina, Institute of Statistics Mimeo Series No. 135, July, 1955, p. 57.

[6] J. W. TUKEY, "Some sampling simplified," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 501–519.

[7] J. W. TUKEY, "Finite sampling simplified," Statistical Research Group, Princeton University, Memorandum Report 45, March,1951.

[8] J. W. TUKEY, "Keeping moment-like sampling computations simple," *Ann. Math. Stat.*, this issue.

[9] J. WISHART, "Moment coefficients of the k-statistics in samples from a finite population," *Biometrika*, Vol. 39 (1952), pp. 1–13.

SOME APPLICATIONS OF BIPOLYKAYS TO THE ESTIMATION OF VARIANCE COMPONENTS AND THEIR MOMENTS[1]

By Robert Hooke[2]

Princeton University

1. Summary. Bipolykays were introduced in [3]. They form a family of symmetric (row-wise and column-wise) polynomial functions of the elements of a two-way array, with the property of being inherited on the average, and such that any similarly symmetric polynomial function of the same numbers can be written linearly in terms of the bipolykays. This paper will describe some applications of bipolykays to problems in the analysis of variance of two-way classifications, using the formulas and tables derived in [3]. A linear model which includes contributions from interaction as well as independently sampled cell contributions is given in Section 3, and applications are made to certain cases of this model. These applications include (a) finding unbiased estimators for the variance components in the case of no interaction as well as unbiased estimators for the variances of these estimators (Section 6), (b) finding expressions for means and variances of some of the functions of degrees 1 and 2 that are of interest in the problem of sampling from a matrix (Section 7), and (c) finding unbiased estimators for variance components in the general case, including expressions for the variances of these estimators in the case of infinite populations (Section 8).

2. Introduction. The purpose of this paper is to describe some uses of bipolykays, which were defined in [3], in connection with problems arising in the analysis of variance. A linear model is given in Section 3 and an analysis of variance notation, in Section 4. Sections 6, 7, and 8 are given over to derivation of results related to the estimation of variance components in various special cases of the linear model.

It will be necessary to make frequent references to [3]. However, in order to enable the reader to get the gist of the present paper without referring to [3], the remainder of this section is devoted to a summary of definitions and frequently-used results.

The symbol $\sum^{\neq}$, for "distinct sum," means a sum taken over all subsequent subscripts, but with subscripts kept different when they are indicated by different letters. Thus

$$\sum_{1}^{2}{}^{\neq} x_i x_j = x_1 x_2 + x_2 x_1 .$$

Received July 28, 1954.

[1] Prepared in connection with research sponsored by the Office of Naval Research.

[2] Now with the Westinghouse Research Laboratories, East Pittsburgh, Pa.

When the x's are matrix elements, the distinctness relates to row subscripts and to column subscripts. Thus

$$\sum_1^{2}{}^{\neq} x_{ij}x_{ik} = x_{11}x_{12} + x_{12}x_{11} + x_{21}x_{22} + x_{22}x_{21}.$$

For a set of numbers $x_1, \cdots, x_n$, the *symmetric means* of degrees 1 and 2 are

$$\langle 1\rangle = \sum{}^{\neq} x_i/n,$$
$$\langle 11\rangle = \sum{}^{\neq} x_ix_j \,/\, n(n-1),$$
$$\langle 2\rangle = \sum{}^{\neq} x_i^2/n.$$

The *polykays* are linear combinations of symmetric means denoted by k's with subscripts. Those of degrees 1 and 2 are defined by

$$k_1 = \langle 1\rangle,$$
$$k_{11} = \langle 11\rangle,$$
$$k_2 = \langle 2\rangle - \langle 11\rangle.$$

Symmetric means and polykays are inherited on the average; this means that if $x_1, \cdots, x_n$ are a sample from a population $x_1, \cdots, x_N$, and if primes are used to denote values defined over the population, then

$$\text{ave } k_2 = k_2', \text{ etc.,}$$

where "ave" means average over all possible samples of size n.

If $x_1, \cdots, x_n$ is a sample from a population P' (with polykays k_1', k_{11}', etc.) and $y_1, \cdots, y_n$ is a sample from a population P'' (with polykays k_1'', k_{11}'', etc.), and if $z_1, \cdots, z_n$ is a sample formed by letting $z_i = x_i + y_i$ $(i = 1, 2, \cdots, n)$, then

$$\text{ave aver } k_1 = k_1' + k_1''. \tag{1}$$

Here k_1 (with no prime) means a polykay over the z's, "aver" means average over all possible permutations (or randomizations) of the x's and y's before adding, and "ave" means average over samples as before. Equation (1) is known as a pairing formula. Pairing formulas for the polykays of degrees 2 are

$$\text{ave aver } k_{11} = k_{11}' + 2k_1'k_1'' + k_{11}'',$$

$$\text{ave aver } k_2 = k_2' + k_2''.$$

We now let $\|x_{ij}\|$ be a two-way array of numbers $(i = 1, 2, \cdots, r;$ $j = 1, 2, \cdots, c)$ which may be regarded as a bisample from an array $\|x_{IJ}\|$ $(I = 1, 2, \cdots, R;\ J = 1, 2, \cdots, C)$, a bisample being chosen from a population matrix by taking those elements which are at the intersections of a selected set of r of the R rows and c of the C columns.

Generalized symmetric means (g.s.m.'s) of degrees 1 and 2 over a bisample are[3]

$$\begin{bmatrix}1 & - \\ - & -\end{bmatrix} = \sum\nolimits^{\neq} x_{ij} \,/\, rc,$$

$$\begin{bmatrix}1 & - \\ - & 1\end{bmatrix} = \sum\nolimits^{\neq} x_{ij}x_{km} \,/\, rc(r-1)(c-1),$$

$$\begin{bmatrix}1 & - \\ 1 & -\end{bmatrix} = \sum\nolimits^{\neq} x_{ij}x_{kj} \,/\, rc(r-1),$$

$$\begin{bmatrix}1 & 1 \\ - & -\end{bmatrix} = \sum\nolimits^{\neq} x_{ij}x_{ik} \,/\, rc(c-1),$$

$$\begin{bmatrix}2 & - \\ - & -\end{bmatrix} = \sum\nolimits^{\neq} x_{ij}^2 \,/\, rc.$$

Those of degrees 3 and 4 can be expressed in similar notation, but to save space are indicated by $t_1, t_2, \cdots, t_{10}$ for the 10 g.s.m.'s of degree 3 and by $f_1, f_2, \cdots, f_{33}$ for the 33 g.s.m.'s of degree 4. An arbitrary symmetric mean may be denoted by $\langle \|\alpha\| \rangle$, the $\|\alpha\|$ representing the matrix of entries.

The bipolykays are linear combinations of the g.s.m.'s, represented in the same way with parentheses replacing $\langle \quad \rangle$'s. Those of degrees 1 and 2 are defined by

$$\begin{pmatrix}1 & - \\ - & -\end{pmatrix} = \begin{bmatrix}1 & - \\ - & -\end{bmatrix},$$

$$\begin{pmatrix}1 & - \\ - & 1\end{pmatrix} = \begin{bmatrix}1 & - \\ - & 1\end{bmatrix},$$

$$\begin{pmatrix}1 & - \\ 1 & -\end{pmatrix} = \begin{bmatrix}1 & - \\ 1 & -\end{bmatrix} - \begin{bmatrix}1 & - \\ - & 1\end{bmatrix},$$

$$\begin{pmatrix}1 & 1 \\ - & -\end{pmatrix} = \begin{bmatrix}1 & 1 \\ - & -\end{bmatrix} - \begin{bmatrix}1 & - \\ - & 1\end{bmatrix},$$

$$\begin{pmatrix}2 & - \\ - & -\end{pmatrix} = \begin{bmatrix}2 & - \\ - & -\end{bmatrix} - \begin{bmatrix}1 & 1 \\ - & -\end{bmatrix} - \begin{bmatrix}1 & - \\ 1 & -\end{bmatrix} + \begin{bmatrix}1 & - \\ - & 1\end{bmatrix}.$$

Those of degrees 3 and 4 are indicated by $T_1, T_2, \cdots, T_{10}$ and $F_1, F_2, \cdots, F_{33}$, respectively. (See Section 8 of [3].) A general bipolykay may be indicated by $(\|\alpha\|)$.

Bipolykays and g.s.m.'s are inherited on the average in the sense that

$$\text{ave}\begin{pmatrix}1 & - \\ - & -\end{pmatrix} = \begin{pmatrix}1 & - \\ - & -\end{pmatrix}', \text{etc.,}$$

[3] Square brackets are used, for convenience in printing, in place of $\langle\rangle$'s for g.s.m.'s having more than one row.

the prime and "ave" having the same meanings as for polykays above, with

$$\begin{pmatrix}1 & -\\ - & -\end{pmatrix}' = \begin{bmatrix}1 & -\\ - & -\end{bmatrix}' = \sum^{\neq} x_{IJ} / RC, \text{etc.}$$

Pairing formulas have a meaning analogous to that defined above for polykays and are as follows for bipolykays of degrees 1 and 2, "aver" meaning average over permutations of rows and of columns:

$$\text{ave aver} \begin{pmatrix}1 & -\\ - & -\end{pmatrix} = \begin{pmatrix}1 & -\\ - & -\end{pmatrix}' + \begin{pmatrix}1 & -\\ - & -\end{pmatrix}'',$$

$$\text{ave aver} \begin{pmatrix}1 & -\\ - & 1\end{pmatrix} = \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}' + 2\begin{pmatrix}1 & -\\ - & -\end{pmatrix}'\begin{pmatrix}1 & -\\ - & -\end{pmatrix}'' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}'',$$

$$\text{ave aver} \begin{pmatrix}1 & -\\ 1 & -\end{pmatrix} = \begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}' + \begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}'',$$

$$\text{ave aver} \begin{pmatrix}1 & 1\\ - & -\end{pmatrix} = \begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' + \begin{pmatrix}1 & 1\\ - & -\end{pmatrix}'',$$

$$\text{ave aver} \begin{pmatrix}2 & -\\ - & -\end{pmatrix} = \begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \begin{pmatrix}2 & -\\ - & -\end{pmatrix}''.$$

In general, it is shown in [3] that the pairing formula for a bipolykay $(\|\alpha\|)$ is

$$\text{ave aver } (\|\alpha\|) = (\|\alpha\|)' + (\|\alpha\|)''$$

unless the matrix $\|\alpha\|$ can be expressed as a direct sum

$$\|\alpha\| = \begin{Vmatrix}\alpha_1 & & 0\\ & \ddots & \\ 0 & & \alpha_m\end{Vmatrix},$$

where the α_i are matrices. If the $\|\alpha_i\|$ cannot be further broken down, then, in this case,

$$\text{ave aver } (\|\alpha\|) = (\|\alpha\|)' + (\|\alpha\|)'' + \sum (\|\beta\|)'(\|\gamma\|)'',$$

where the sum extends over all $\|\beta\|$ and $\|\gamma\|$ such that $\|\beta\|$ is the direct sum of $1, 2, \cdots$, or $m - 1$ of the $\|\alpha_i\|$ and $\|\gamma\|$ is the direct sum of the remaining ones.

The reader is referred to [3] for the general definitions exemplified above, for multiplication formulas, for conversion formulas for degrees 3 and 4, etc.

3. The linear model. Each example discussed in this paper will be based on a linear model which is a special case of the following:

$$x_{ijk} = \theta + \eta_i + \xi_j + \lambda_{ij} + \omega_{ijk}. \tag{2}$$

Here i, j, k run from 1 to r, c, b, respectively. The θ's, η's, ξ's, and ω's are independently sampled contributions from populations described in Table 1.

The systematic interactions λ_{ij} are not independently sampled but are "tied" to the η's and ξ's; i.e., the λ's come from an $R \times C$ matrix having a row cor-

TABLE 1

Notation associated with model (2)

Contribution	Sample Size	Population Size	Population Polykays
θ, general	1	Arbitrary	k_1'''', k_{11}'''', etc.
η_i, row	r	R	k_1', k_{11}', etc. ($k_1' = 0$)
ξ_j, column	c	C	k_1'', k_{11}'', etc. ($k_1'' = 0$)
ω_{ijk}, cell	rcb	N	k_1''', k_{11}''', etc. ($k_1''' = 0$)

responding to each η and a column to each ξ, so that for each selection of an η_i and a ξ_j there is a unique λ_{ij} that accompanies them. Since the η's and ξ's represent row and column contributions, respectively, it is assumed that the row means and column means of the matrix of λ's all vanish; otherwise the population matrix of λ's is arbitrary and so must be described in terms of bipolykays rather than polykays.

4. Analysis of variance notation. The matrix

$$\|x_{ij}\|, \quad i = 1, 2, \cdots, r; j = 1, 2, \cdots, c$$

will represent a bisample from a population matrix

$$\|x_{IJ}\|, \quad I = 1, 2, \cdots, R; J = 1, 2, \cdots, C.$$

For any matrix $\|x_{ij}\|$, whether it be a population matrix, a bisample from such a matrix, or simply a two-way array of sampled numbers such as arises in connection with some linear model, our interests will center around certain families of symmetric quadratic functions of the x's. Two of these are

(a) the bipolykays $\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}, \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}, \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$, and

(b) the various sums of squares and mean squares associated with conventional analysis of variance procedures.

The mean squares (denoted by MS) and sums of squares (denoted by SS) are defined as follows, where a dot represents an average over the subscript it replaces:

Designation	*Mean Square and Sum of Squares*
Rows	$MSR = SSR/(r-1) = c \sum_i (x_{i.} - x_{..})^2/(r-1)$
Columns	$MSC = SSC/(c-1) = r \sum_j (x_{.j} - x_{..})^2/(c-1)$
Residual or balance	$MSB = SSB/(r-1)(c-1)$
	$= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2/(r-1)(c-1)$
Mean	$MSM = SSM = \sum_i \sum_j x_{..}^2 = rcx_{..}^2$
Total	$SST = \sum_i \sum_j (x_{ij} - x_{..})^2$

(As always, we have $SST = SSR + SSC + SSB$.)

When it is desired to emphasize the fact that a population matrix is being

discussed, subscripts will be capital letters, dashes (instead of dots) will indicate averages over the population, and primes will indicate population values; e.g.,

$$SST' = \sum_{I=1}^{R} \sum_{J=1}^{C} (x_{IJ} - x_{--})^2.$$

In dealing with a population matrix we shall use the following quantities:

$$\theta = x_{--} \qquad \text{(i.e., } \theta = \sum \sum x_{IJ}/RC),$$

$$\eta_I = x_{I-} - x_{--},$$

$$\xi_J = x_{-J} - x_{--},$$

$$\lambda_{IJ} = x_{IJ} - x_{I-} - x_{-J} + x_{--}.$$

The elements of the matrix can then be thought of as built up from the η's, ξ's, λ's, and θ as in model (2), taking the ω's in that model to be 0. We are then interested in a third family of quadratic symmetric functions, namely

(c) the variance components,

$$\begin{aligned} \sigma_R^2 &= \text{variance component for rows} \\ &= \textstyle\sum_I (x_{I-} - x_{--})^2 / (R-1) \\ &= k_2', \end{aligned}$$

where k_2' has the meaning assigned in Table 1;

$$\begin{aligned} \sigma_C^2 &= \text{variance component for columns} \\ &= \textstyle\sum_J (x_{-J} - x_{--})^2 / (C-1) \\ &= k_2''; \end{aligned}$$

$$\sigma_\lambda^2 = \textstyle\sum_I \sum_J (x_{IJ} - x_{I-} - x_{-J} + x_{--})^2 / (R-1)(C-1).$$

In this section we shall express the SS's in terms of the bipolykays (so that moments of the former can be easily obtained), the bipolykays in terms of the MS's (so that values of the former can be computed by standard techniques used in computing the latter), and finally the MS's for a population matrix in terms of the variance components (to help in expressing unbiased estimators for the latter). These various expressions are derived by elementary algebra, so only one example of the derivations will be given:

To express SSR in terms of bipolykays, for example, we have

$$\begin{aligned} SSR &= c \textstyle\sum_i (x_{i.} - x_{..})^2 \\ &= c \textstyle\sum_i x_{i.}^2 - rcx_{..}^2. \end{aligned}$$

Since

$$\begin{aligned} c \textstyle\sum_i x_{i.}^2 &= c \textstyle\sum_i \left(\sum_j x_{ij} / c\right)^2 \\ &= \frac{1}{c} \textstyle\sum_i \left(\sum_j x_{ij}^2 + \sum_{j,k}^{\neq} x_{ij} x_{ik}\right) \\ &= \frac{1}{c}\left\{ rc \begin{bmatrix} 2 & - \\ - & - \end{bmatrix} + rc(c-1) \begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix} \right\}, \end{aligned}$$

and since

$$
\begin{aligned}
rc\bar{x}_{..}^2 &= rc\left(\textstyle\sum_{i,j} x_{ij} / rc\right)^2 \\
&= \frac{1}{rc}\left(\textstyle\sum_{i,j} x_{ij}^2 + \sum^{\neq} x_{ij}x_{ik} + \sum^{\neq} x_{ij}x_{kj} + \sum^{\neq} x_{ij}x_{km}\right) \\
&= \frac{1}{rc}\left\{rc\begin{bmatrix} 2 & - \\ - & - \end{bmatrix} + rc(c-1)\begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix} + rc(r-1)\begin{bmatrix} 1 & - \\ 1 & - \end{bmatrix}\right. \\
&\qquad\qquad \left. + rc(r-1)(c-1)\begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}\right\},
\end{aligned}
$$

we have

$$
SSR = (r-1)\left\{\begin{bmatrix} 2 & - \\ - & - \end{bmatrix} - \begin{bmatrix} 1 & - \\ 1 & - \end{bmatrix} + (c-1)\left(\begin{bmatrix} 1 & 1 \\ - & - \end{bmatrix} - \begin{bmatrix} 1 & - \\ - & 1 \end{bmatrix}\right)\right\}.
$$

The equations defining the bipolykays of degrees 1 and 2 (Section 2) enable us to express SSR at once as a linear combination of bipolykays. The result, together with similar ones for the other SS's, is contained in Table 2.

The equations represented by the first four rows of Table 2 can be solved to produce the following inverse relationships, where it is convenient to use MS's in place of SS's:

$$\begin{pmatrix} 2 & - \\ - & - \end{pmatrix} = MSB, \tag{3}$$

$$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} = (MSR - MSB) / c, \tag{4}$$

$$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix} = (MSC - MSB) / r, \tag{5}$$

$$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix} = (MSM - MSR - MSC + MSB) / rc. \tag{6}$$

TABLE 2

Coefficients for SS's as linear combinations of bipolykays

	$\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}$	$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}$	$\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}$
SSR	$r-1$	$c(r-1)$	0	0
SSC	$c-1$	0	$r(c-1)$	0
SSB	$(r-1)(c-1)$	0	0	0
SSM	1	c	r	rc
SST	$rc-1$	$c(r-1)$	$r(c-1)$	0

Finally, it follows immediately from the definitions that, for a population matrix,

$$\sigma_R^2 = MSR' \,/\, C$$

(7)
$$= \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}' + \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}' \Big/ C,$$

$$\sigma_C^2 = MSC' \,/\, R$$

(8)
$$= \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}' + \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}' \Big/ R,$$

$$\sigma_\lambda^2 = MSB'$$

(9)
$$= \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}'.$$

5. Expectations in the case of no interactions. We suppose now that all the λ's in (2) are zero, and for the present that $b = 1$, so that the model is

(10) $$x_{ij} = \theta + \eta_i + \xi_j + \omega_{ij}, \qquad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c.$$

Our first problem is to find the average values of bipolykays defined over the $r \times c$ array of x's (and distinguished here by having no prime or asterisk) in terms of the polykays (one to four primes) of the four populations described in Table 1. The procedure is to apply bipolykay pairing formulas to (10). The populations from which come the η's, ξ's, ω's, and θ are all special cases (numbered I to IV, respectively, below) and will now be considered one at a time.

CASE I. The population of η's is not a matrix population, but it can be thought of as a matrix whose Ith row is a vector of C components, each equal to η_I, $(I = 1, 2, \cdots, R)$. Any sample of r η's can then be regarded as a bisample of r rows and c (arbitrary) columns from this matrix. Bipolykays of this bisample will be written $(\| \alpha \|)^*$, with a single asterisk. Referring to Case I of Section 7 of [3], we see that

$$(\| \alpha \|)^* = k^*_{mn\cdots p} \qquad \text{if the entries of } (\| \alpha \|)^* \text{ are all 1's in different columns}$$

$$= 0 \qquad \text{otherwise},$$

$m, n, \cdots p$ being the row sums of the entries in $\| \alpha \|$, and k^* denoting a polykay for the sample $\eta_1, \cdots, \eta_r$ which defined the bisample in question. The same is obviously true for the population.

CASE II. The remarks just made for the η's apply to the population of ξ's if we change rows to columns, single primes and asterisks to double primes and asterisks, etc.

CASE III. The population of ω's enters as in Case IV of Section 7 of [3]. There

it was shown that $(\| \alpha \|)^{***}$ becomes, on the average (over the kind of randomization that is pertinent here),

$k^{***}_{mn\cdots p}$	if $m, n, \cdots, p$ are the entries of $\| \alpha \|$ and if all are in different row and in different columns
0	otherwise.

CASE IV. In sampling θ we take a sample of size 1 and make it a bisample by putting this one number into every cell of the $r \times c$ matrix. Referring to Case III of Section 7 of [3], we see that

$(\| \alpha \|)^{****} = \theta^m$	if all m entries of $\| \alpha \|$ are 1's in different rows and different columns
$= 0$	otherwise

Hence ave $(\| \alpha \|)^{****} = \langle m \rangle''''$ or 0, respectively.

Keeping these facts in mind, together with the fact that $k_1' = k_1'' = k_1''' = 0$, we can apply the pairing formulas to the bipolykays of degree 4 or less to obtain some useful results that are collected in Table 3 below. We first derive a few of these results to show how the pairing formulas are used.

The only first-degree bipolykay is, of course, indecomposable, so its pairing formula (Section 2) gives us

$$\begin{aligned}\text{ave aver}\begin{pmatrix}1 & -\\ - & -\end{pmatrix} &= \begin{pmatrix}1 & -\\ - & -\end{pmatrix}' + \begin{pmatrix}1 & -\\ - & -\end{pmatrix}'' + \begin{pmatrix}1 & -\\ - & -\end{pmatrix}''' + \begin{pmatrix}1 & -\\ - & -\end{pmatrix}''''\\ &= k_1' + k_1'' + k_1''' + k_1''''\\ &= k_1'''',\end{aligned}$$

since $k_1' = k_1'' = k_1''' = 0$.

The indecomposable bipolykays of degree 2 can be treated in exactly the same way; *e.g.*

$$\begin{aligned}\text{ave aver}\begin{pmatrix}2 & -\\ - & -\end{pmatrix} &= \begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \begin{pmatrix}2 & -\\ - & -\end{pmatrix}'' + \begin{pmatrix}2 & -\\ - & -\end{pmatrix}''' + \begin{pmatrix}2 & -\\ - & -\end{pmatrix}''''\\ &= k_2''',\end{aligned}$$

since the term $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}'''$ is really aver $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}^{***}$, which is k_2''' by the remarks under Case III above. The other terms, i.e. those with one, two, and four primes, vanish in accordance with the remarks made in Cases I, II, and IV, above.

Decomposable bipolykays lead to more complex expressions. Averaging $\begin{pmatrix}1 & -\\ - & 1\end{pmatrix}$, for example, produces

$$\begin{aligned}\text{ave aver}\begin{pmatrix}1 & -\\ - & 1\end{pmatrix} &= \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}'' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}''' + \begin{pmatrix}1 & -\\ - & 1\end{pmatrix}''''\\ &\qquad + 2\begin{pmatrix}1 & -\\ - & -\end{pmatrix}'\begin{pmatrix}1 & -\\ - & -\end{pmatrix}'' + \cdots.\\ &= k_{11}' + k_{11}'' + k_{11}''' + \langle 2 \rangle''''.\end{aligned}$$

As an example for degree 4, we consider the decomposable bipolykay

$$F_{11} = \begin{pmatrix} - & 1 & 1 \\ 1 & - & - \\ 1 & - & - \end{pmatrix}.$$

The pairing formula is

$$\text{ave aver } F_{11} = \text{ave aver } (F_{11}^{*} + F_{11}^{**} + F_{11}^{***} + F_{11}^{****})$$

$$+ \text{ ave aver } \sum \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^{u} \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}^{v},$$

where u and v represent different numbers (1, 2, 3, or 4) of asterisks. By the remarks made under Cases I through IV, each term of the form ave aver F_{11}^{*} vanishes, as does each term ave aver $\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^{u}$ ave aver $\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}^{v}$ except in the case where u and v represent a single and double asterisk, respectively. (The assumption of independence in sampling provides that the average of the sum of products is equal to the sum of products of averages.) Hence we have

$$\begin{aligned} \text{ave aver } F_{11} &= \text{ave aver } \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^{*} \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}^{**} \\ &= k_2' k_2'' \end{aligned}$$

Continuing in this way, we obtain the results shown in Table 3.

The following, omitted from Table 3, have expectation 0: T_6, T_8, T_9; F_{10}, F_{13}, F_{14}, F_{15}, F_{16}, F_{17}, F_{20}, F_{21}, F_{22}, F_{23}, F_{24}, F_{25}, F_{26}, F_{29}, F_{30}, F_{31}, and F_{32}.

The following have expectations which are complex expressions that will not be used in this paper: $\begin{pmatrix} 1 & - \\ - & 1 \end{pmatrix}$; T_1, T_2, T_3, T_7; F_1, F_2, F_3, F_6, F_7, F_{12}, and F_{28}.

The first formula in Table 3 says that $\begin{pmatrix} 1 & - \\ - & - \end{pmatrix}$, the mean of the x's, is an un-

TABLE 3

Bipolykays (column A) and their expected values (column B) in model (10)

A	B	A	B
$\begin{pmatrix} 1 & - \\ - & - \end{pmatrix}$	k_1''''	F_4	k_{22}'
		F_5	k_{22}''
$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}$	k_2'	F_8	k_4'
		F_9	k_4''
$\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}$	k_2''	F_{11}	$k_2'k_2''$
		F_{18}	$k_2'k_2'''$
$\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$	k_2'''	F_{19}	$k_2''k_2'''$
		F_{27}	k_{22}'''
T_4	k_3'	F_{28}	k_4'''
T_5	k_3''		
T_{10}	k_3'''		

biased estimator of the mean of the θ population, which is obvious, since the other populations have zero means. The second formula (reading down) says that $\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}$ is an unbiased estimator of the component of variance for rows. In (4) it was pointed out that

$$\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} = (MSR - MSB) / c,$$

so this is the usual result. Similarly for the third formula. The fourth says that $\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$ is an unbiased estimator of the "error variance." Since $\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}$ is MSB, this is well known. Interpretation of the formulas for the F's will be given in Section 6.

The two-way model with replications,

$$x_{ijk} = \theta + \eta_i + \xi_j + \omega_{ijk}, \qquad k = 1, 2, \cdots, b, \tag{11}$$

where everything is as in (10) except for the greater number of ω's sampled, can be treated in the same way if we suppose the population of ω's to be infinite. First we find $x_{ij} = x_{ij\cdot}$, the average in each cell. Then

$$x_{ij} = \theta + \eta_i + \xi_j + \omega_{ij\cdot},$$

and we have the same situation as before, with this exception: if P_1 represents the population of ω's, the $\omega_{ij\cdot}$ come from the population P_2 of samples of size b from P_1. It remains only to find the polykays of P_2 in terms of those of P_1.

Any $\omega_{ij+} = \sum_k \omega_{ijk}$ can be thought of as a sum of b numbers from b populations all equal to P_1. Hence if $k_{22}^{***(+)}$ $k_{22}^{***(\cdot)}$ represent polykays for the ω_{ij+} and $\omega_{ij\cdot}$, respectively, we have, by the pairing formula for k_{22}, for example,

$$\begin{aligned} \text{ave aver } k_{22}^{***(+)} &= bk_{22}''' + b(b-1)k_2'''k_2''' \\ &= b^2 k_{22}'''. \end{aligned}$$

Finally, since k_{22} is of degree 4, we divide by b^4 to obtain

$$\text{ave aver } k_{22}^{***(\cdot)} = k_2'''/b^2.$$

In similar fashion we find that

$$\begin{aligned} \text{ave aver } k_2^{***(\cdot)} &= k_2'''/b, \\ \text{ave aver } k_3^{***(\cdot)} &= k_3'''/b^2, \\ \text{ave aver } k_4^{***(\cdot)} &= k_4'''/b^3. \end{aligned}$$

It follows that for model (11), Table 3 remains as it is, except that k_2''', k_3''', k_{22}''', and k_4''', wherever they appear, must be divided, respectively, by b, b^2, b^2, and b^3. For example,

$$\text{aver aver } F_{18} = k_2'k_2'''/b,$$

F_{18} being defined over the matrix $\|x_{ij\cdot}\|$.

6. Estimating variance components and their variances in the case of no interactions. In this section we consider applications to the case described by model (10) or (11). We begin with model (10), where components of variance for rows, columns, and "error" are k_2', k_2'', and k_2''', respectively, and for which respective unbiased estimators (Table 3) are $\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}$, $\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}$, and $\begin{pmatrix}2 & -\\ - & -\end{pmatrix}$. In order to find the variances or covariances of these estimates, one may proceed as in this example:

$$\text{(12)}\qquad \begin{aligned}\text{var}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix} &= \text{ave aver}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}^2 - \left\{\text{ave aver}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}\right\}^2\\ &= \text{ave aver}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}^2 - (k_2')^2.\end{aligned}$$

Referring to the relevant multiplication formulas, we find that

$$(k_2')^2 = \frac{R+1}{R-1}k_{22}' + \frac{1}{R}k_4'$$

([5], p. 516) and that

$$\begin{aligned}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}^2 = [rc(r+1)(c-1)F_4 &+ c(r-1)(c-1)F_8\\ &+ 4r^2(c-1)F_{13} + 2r^2F_{17} + 4r(c-1)F_{18}\\ &+ 4(r-1)(c-1)F_{22} + 2rF_{27}\\ &+ 2(r-1)F_{29}] / rc(r-1)(c-1)\end{aligned}$$

by Section 9 of [3]).

From Table 3 it follows that

$$\begin{aligned}\text{ave aver}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}^2 = [rc(r+1)(c-1)k_{22}' &+ c(r-1)(c-1)k_4' + 4r(c-1)k_2'k_2'''\\ &+ 2rk_{22}'''] / rc(r-1)(c-1).\end{aligned}$$

Hence (12) becomes

$$\text{(13)}\qquad \begin{aligned}\text{var}\begin{pmatrix}1 & 1\\ - & -\end{pmatrix} = \frac{2}{c(r-1)(c-1)}k_{22}''' &+ \frac{4}{c(r-1)}k_2'k_2'''\\ &+ \left(\frac{2}{r-1} - \frac{2}{R-1}\right)k_{22}' + \left(\frac{1}{r} - \frac{1}{R}\right)k_4'.\end{aligned}$$

In similar fashion one obtains

$$\text{(14)}\qquad \begin{aligned}\text{var}\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix} = \frac{2}{r(r-1)(c-1)}k_{22}''' &+ \frac{4}{r(c-1)}k_2''k_2''',\\ &+ \left(\frac{2}{c-1} - \frac{2}{C-1}\right)k_{22}'' + \left(\frac{1}{c} - \frac{1}{C}\right)k_4'',\end{aligned}$$

(15)
$$\operatorname{var}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix} = \left(\frac{1}{rc} - \frac{1}{N}\right) k_4''' + \left\{\frac{2}{(r-1)(c-1)} - \frac{2}{(N-1)}\right\} k_{22}''',$$

(16)
$$\operatorname{cov}\left\{\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}, \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}\right\} = 2k_{22}''' \,/\, rc(r-1)(c-1),$$

(17)
$$\operatorname{cov}\left\{\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}, \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}\right\} = -2k_{22}''' \,/\, c(r-1)(c-1),$$

(18)
$$\operatorname{cov}\left\{\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}, \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}\right\} = -2k_{22}''' \,/\, r(r-1)(c-1).$$

Formulas (13) through (18) have, of course, been derived without the use of bipolykays, and are now new. Our interest here is in using them to derive unbiased estimators of the variances and covariances of estimated variance components. In the above formulas each variance is given in terms of population polykays k_2', k_{22}'', etc. If one could observe the actual η's, ξ's, and ω's that are sampled, the sample polykays k_2^*, k_{22}^{**}, etc., would then provide unbiased estimators of the population polykays. In practice, however, one cannot do this, so that the formulas above do not provide unbiased estimates.

Inspection of formulas (13) through (18) shows that one would like to have unbiased estimators for the following:

$$k_{22}', \; k_{22}'', \; k_{22}'''; \; k_2'k_2''', \; k_2''k_2'''; \; k_4', \; k_4'', \; k_4'''.$$

Such estimators are provided at once by Table 3, and are, respectively, the following bipolykays computed over the matrix of x's:

$$F_4, \; F_5, \; F_{27}, \; F_{18}, \; F_{19}, \; F_8, \; F_9, \; F_{33}.$$

Substituting these into formula (13), we have

$$\text{Unbiased estimator for } \operatorname{var}\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} = \frac{2}{c(r-1)(c-1)} F_{27} + \frac{4}{c(r-1)} F_{18} + \left(\frac{2}{r-1} - \frac{2}{R-1}\right) F_4 + \left(\frac{1}{r} - \frac{1}{R}\right) F_8,$$

and similarly for the other formulas.

Turning now to model (11), with replications, and again supposing the population of ω's to be infinite, we note that estimators for k_2', k_2'', and k_2''' are, respectively, $\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}$, $\begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix}$, and $\begin{pmatrix} 2 & - \\ - & - \end{pmatrix} \Big/ b$, these bipolykays being applied to the matrix $\|x_{ij.}\|$. Hence (13) through (18) give us the formulas for the variances and covariances of these estimators if only we divide the entire right-hand sides of (15), (17), and (18) by b^2, b, and b, respectively, and change (in all for-

mulas) all of the three-primed quantities in the manner indicated at the end of Section 5.

7. Moments in bisampling. The problem that led originally to the development of bipolykays (see [1]) came from this model of the educational testing process: Given a population of C questions (a "test") and a population of R examinees, suppose that the score of examinee I on question J will be x_{IJ}. A "test form," consisting of a random sample of c of the C questions, is given to each of a random sample of r of the R examinees. The "test score" of the ith examinee is $\sum_{j=1}^{c} x_{ij}$, the average test score of the group is $\sum_i \sum_j x_{ij}/r$, etc. One wants means and variances of quantities such as these.

Insofar as problems connected with bisampling can be expressed in terms of finding low moments of first- and second-degree symmetric polynomial functions, they can be easily solved by application of bipolykays. For degree 1, we have, for example,

$$\operatorname{var}\begin{pmatrix}1 & -\\ - & -\end{pmatrix} = \left(\frac{1}{rc} - \frac{1}{RC}\right)\begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + \left(\frac{1}{r} - \frac{1}{R}\right)\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' + \left(\frac{1}{c} - \frac{1}{C}\right)\begin{pmatrix}1 & -\\ 1 & -\end{pmatrix}' \qquad \text{by Section 10 of [3]}$$

$$= \frac{1}{c}\left(\frac{1}{r} - \frac{1}{R}\right) MSR' + \frac{1}{r}\left(\frac{1}{c} - \frac{1}{C}\right) MSC' - \left(\frac{1}{c} - \frac{1}{C}\right)\left(\frac{1}{r} - \frac{1}{R}\right) MSB' \qquad \text{by (3), (4), (5).}$$

As far as first moments of functions of degree 2 are concerned, we can derive the following formulas:

$$\begin{aligned} E(MSR) &= MSB' + \frac{c}{C}(MSR' - MSB') \\ &= \left(1 - \frac{c}{C}\right)\sigma_\lambda^2 + c\sigma_R^2. \end{aligned} \tag{19}$$

$$\begin{aligned} E(MSC) &= MSB' + \frac{r}{R}(MSC' - MSB') \\ &= \left(1 - \frac{r}{R}\right)\sigma_\lambda^2 + r\sigma_C^2. \end{aligned} \tag{20}$$

$$E(MSB) = \sigma_\lambda^2. \tag{21}$$

We use one of these to illustrate the derivations:

$$(ESMR) = \operatorname{ave}\left\{\begin{pmatrix}2 & -\\ - & -\end{pmatrix} + c\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}\right\} \qquad \text{from Table 2}$$

$$= \begin{pmatrix}2 & -\\ - & -\end{pmatrix}' + c\begin{pmatrix}1 & 1\\ - & -\end{pmatrix}' \qquad \text{by inheritance on the average}$$

$$= MSB' + \frac{c}{C}(MSR' - MSB') \qquad \text{by (3) and (4)}$$

$$= \left(1 - \frac{c}{C}\right)\sigma_\lambda^2 + c\sigma_R^2 \qquad \text{by (7) and (9).}$$

Equations (19), (20), and (21) lead at once to the following unbiased estimators of the σ^2's:

$$\hat{\sigma}_\lambda^2 = MSB = \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}, \tag{22}$$

$$\hat{\sigma}_R^2 = \frac{1}{c}MSR - \left(\frac{1}{c} - \frac{1}{C}\right)MSB = \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} + \begin{pmatrix} 2 & - \\ - & - \end{pmatrix} \Big/ C, \tag{23}$$

$$\hat{\sigma}_C^2 = \frac{1}{r}MSC - \left(\frac{1}{r} - \frac{1}{R}\right)MSB = \begin{pmatrix} 1 & - \\ 1 & - \end{pmatrix} + \begin{pmatrix} 2 & - \\ - & - \end{pmatrix} \Big/ R. \tag{24}$$

We turn next to the variances of the functions of degree 2. Variances and covariances of the bipolykays were tabulated in Section 10 of [3]. From these it is easy to find expressions, in terms of the bipolykays of degree 4, for the variances of the mean squares of the bisample, though these expressions are quite long. To find the variance of MSR, for example, we recall from Table 2 that

$$MSR = \begin{pmatrix} 2 & - \\ - & - \end{pmatrix} + c\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}.$$

Hence

$$\operatorname{var} MSR = \operatorname{var}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix} + 2c \operatorname{cov}\left\{\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}\right\} + c^2 \operatorname{var}\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}.$$

The three terms on the right-hand side of this equation are given in Section 10 of [3], leading to an expression for var MSR in terms of bipolykays of degree 4. Variances of the estimated variance components can be found in the same way. These expressions for variances of quadratic expressions are long and clumsy. Formulas for unbiased estimators of these variances, however, are less complicated. Suppose, for example, that we want an unbiased estimator for the variance of $\hat{\sigma}_R^2$. We have

$$\begin{aligned}
\operatorname{var} \hat{\sigma}_R^2 &= \operatorname{ave} (\hat{\sigma}_R^2)^2 - (\operatorname{ave} \hat{\sigma}_R^2)^2 \\
&= \operatorname{ave} \hat{\sigma}_R^4 - \left\{\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}' + \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}' \Big/ C\right\}^2 && \text{by (7)} \\
&= \operatorname{ave} \hat{\sigma}_R^4 - \left\{\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}'^2 + 2C\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}'\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}' + C^2\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}'^2\right\} \Big/ C^2 \\
&= \operatorname{ave} \hat{\sigma}_R^4 - \frac{1}{C^3}\left\{\frac{R+1}{R-1}[F'_{30} + C(2F'_{17} + 4F'_{25} + F'_{27}) + C^2(4F'_{12} + 2F'_{18}) \right. \\
&\qquad \left. + C^3F'_4] + \frac{1}{R}[F'_{33} + C(3F'_{29} + 4F'_{31}) + 6C^2F'_{22} + C^3F'_8]\right\},
\end{aligned}$$

the last step following from the multiplication table in Section 9 of [3]. It follows that we have this unbiased estimator of var $\hat{\sigma}_R^2$:

$$\hat{\sigma}_R^4 - \frac{1}{C^3}\left\{\frac{R+1}{R-1}\,[F_{30} + C(2F_{17} + 4F_{25} + F_{27}) + C^2(4F_{13} + 2F_{18}) + C^3F_4] + \frac{1}{R}\,[F_{33} + C(3F_{29} + 4F_{31}) + 6C^2F_{22} + C^3F_8]\right\}.$$

In a numerical case, an estimator for var $\hat{\sigma}_R^2$ is not likely to be wanted unless $\hat{\sigma}_R^2$ itself has already been computed, so there is no reason for expanding $\hat{\sigma}_R^4$ in bipolykays. It is of interest to note that the exponent of C (inside square brackets above) is one less than the number of columns appearing in the primary notation for each of the accompanying bipolykays; those bipolykays occurring in the first set of square brackets each have two rows, and those in the second set have one row.

8. Analysis of variance with interaction. We return now to the full model (2), supposing only for the present that $b = 1$, so that the model is

$$x_{ij} = \theta + \eta_i + \xi_j + \lambda_{ij} + \omega_{ij}\,. \tag{25}$$

That part of the sampling which pertains to the η's, ξ's, λ's, and θ can be described more simply (from the algebraic point of view, at least) by saying that

$$\rho_{ij} = \theta + \eta_i + \xi_j + \lambda_{ij}\,, \qquad i = 1, 2, \cdots, r; j = 1, 2, \cdots, c,$$

form a bisample from some $R \times C$ matrix. (See Section 4.) The model (25) then reduces to

$$x_{ij} = \rho_{ij} + \omega_{ij}\,, \tag{26}$$

where the ρ_{ij} are a bisample from a matrix $\| \rho_{IJ} \|$ for which σ_R^2, σ_C^2, and σ_λ^2 are defined as in Section 4, and the ω's are a sample of size rc from a population of variance σ^2. We shall use here one and two primes or asterisks to refer to functions over populations or bisamples related to ρ and ω, respectively.

To find $E(MSR)$, say, for this model, we recall that

$$MSR = \begin{pmatrix} 2 & - \\ - & - \end{pmatrix} + c\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix} \qquad \text{from Table 2.}$$

Taking averages and using the pairing formulas (Section 2), we have

$$E(MSR) = \text{ave aver}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}^* + \text{ave aver}\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}^{**} + c\left\{\text{ave aver}\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^* + \text{ave aver}\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^{**}\right\}.$$

For the population of ω's we have aver $\begin{pmatrix} 2 & - \\ - & - \end{pmatrix}^{**} = k_2''$ and aver $\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}^{**} = 0,$

as was pointed out in Section 5. Hence

$$E(MSR) = \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}' + k'' + c\begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}'$$

$$= \left(1 - \frac{c}{C}\right)\sigma_\lambda^2 + c\sigma_R^2 + \sigma^2 \qquad \text{by (3) and (5).}$$

Similarly,

$$E(MSC) = \left(1 - \frac{r}{R}\right)\sigma_\lambda^2 + r\sigma_C^2 + \sigma^2$$

and

$$E(MSB) = \sigma_\lambda^2 + \sigma^2.$$

Such results have been obtained before, for example in [4], though perhaps not so simply.

In model (26) the lack of replication of course leaves the ω's and λ's confounded. We suppose now that there are b observations per cell, so that the model can be written

$$x_{ijk} = \rho_{ij} + \omega_{ijk}, \qquad k = 1, 2, \cdots, b. \tag{27}$$

As in section 5, we consider the matrix of

$$x_{ij.} = \rho_{ij} + \omega_{ij.},$$

supposing again that the population of ω's is infinite. We then obtain

$$E(\overline{MSR}) = \left(1 - \frac{c}{C}\right)\sigma_\lambda^2 + c\sigma_R^2 + \sigma^2/b, \tag{28}$$

$$E(\overline{MSC}) = \left(1 - \frac{r}{R}\right)\sigma_\lambda^2 + r\sigma_C^2 + \sigma^2/b, \tag{29}$$

$$E(\overline{MSB}) = \sigma_\lambda^2 + \sigma^2/b, \tag{30}$$

where $\overline{MSR}$ means MSR for the matrix $\| x_{ij.} \|$, etc. Since

$$MSW = \frac{1}{rc}\sum_{i,j}\frac{1}{b-1}\sum_k (x_{ijk} - x_{ij.})^2$$

has expectation σ^2, we find from (28), (29), and (30) the following unbiased estimators for the variance components:

$$\hat{\sigma}^2 = MSW,$$

$$\hat{\sigma}_\lambda^2 = \overline{MSB} - MSW/b,$$

$$\hat{\sigma}_R^2 = \overline{MSR}/c - \left(\frac{1}{c} - \frac{1}{C}\right)\overline{MSB} - MSW/bC,$$

$$\hat{\sigma}_C^2 = \overline{MSC}/r - \left(\frac{1}{r} - \frac{1}{R}\right)\overline{MSB} - MSW/bR.$$

One would like to know the variances of these estimators, but in general the third dimension introduced by the subscript k seems to preclude finding them by means of bipolykays. However, if we are willing to assume that R and C are infinite, we have

$$\hat{\sigma}_R^2 = (\overline{MSR} - \overline{MSB})/c = \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix},$$

$$\hat{\sigma}_C^2 = (\overline{MSC} - \overline{MSB})/r = \begin{pmatrix} \overline{1\ -} \\ 1\ - \end{pmatrix},$$

and the variances of these can be found as follows:

$$\begin{aligned}\operatorname{var} \hat{\sigma}_R^2 &= \text{ave aver} \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix}^2 - \left\{ \text{ave aver} \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix} \right\}^2 \\ &= \text{ave aver} \sum a_i \bar{F}_i - \left\{ \text{ave aver} \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix} \right\}^2,\end{aligned}$$

where $\sum a_i \bar{F}_1$ is the expression for $\begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix}^2$ in terms of bipolykays of degree 4, the a_i being functions of r and c given in Section 8 of [3]. In this sum 8 distinct F's appear, and their pairing formulas in the present situation become as follows, by virtue of the remarks under Case III in Section 5:

$$\text{ave aver } \bar{F}_{27} = F_{27}' + 2 \begin{pmatrix} 2\ - \\ -\ - \end{pmatrix}' \bar{k}_2'' + \bar{k}_{22}'',$$

$$\text{ave aver } \bar{F}_{18} = F_{18}' + \begin{pmatrix} 1\ 1 \\ -\ - \end{pmatrix}' \bar{k}_2'',$$

$$\text{ave aver } \bar{F}_i = F_i', \qquad i = 4, 8, 13, 17, 22, 29,$$

where $\bar{k}_2''$, etc., are polykays of the population of $\omega_{ij.}$ and can be expressed in terms of the polykays of the ω_{ijk} as at the end of Section 5. Finally we have, again from Case III in Section 5,

$$\text{ave aver} \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix} = \begin{pmatrix} 1\ 1 \\ -\ - \end{pmatrix}',$$

and so

$$\left\{ \text{ave aver} \begin{pmatrix} \overline{1\ 1} \\ -\ - \end{pmatrix} \right\}^2 = F_4',$$

by Section 9 of [3], since R and C are infinite. Hence

$$\begin{aligned} rc(r-1)(c-1) \operatorname{var} \hat{\sigma}_R^2 = {} & 2rc(c-1)F_4' + c(r-1)(c-1)F_8' \\ & + 4r^2(c-1)F_{13}' + 2r^2F_{17}' + 4r(c-1)F_{18}' \\ & + 4(r-1)(c-1)F_{22}' + 2rF_{27}' + 2(r-1)F_{29}' \end{aligned}$$

$$+ 4r(c - 1) \begin{pmatrix} 1 & 1 \\ - & - \end{pmatrix}' k_2''/b$$

$$+ 4r \begin{pmatrix} 2 & - \\ - & - \end{pmatrix}' k_2''/b + 2rk_2''/b^2.$$

9. Computation. In order to make use of the formulas developed in this paper, it is of course necessary to be able to compute the bipolykays in particular numerical situations. Those of degree 2 can be easily found from equations (3) through (6) of Section 4, after *MSR*, *MSC*, etc., have been computed by standard procedures. A method of computation has been developed for the bipolykays of degree 4, but it will not be given here, as it is very lengthy, and it is hoped that better procedures can be found; this method was reported in [2], copies of which may be obtained on request by writing the secretary of the Statistical Research Group, Box 708, Princeton, N. J.

REFERENCES

[1] R. Hooke, "Sampling from a matrix, with applications to the theory of testing," Statistical Research Group, Princeton University, Memorandum Report 53, November, 1953.

[2] R. Hooke, "The estimation of polykays in the analysis of variance," Statistical Research Group, Princeton University, Memorandum Report 56, May, 1954.

[3] R. Hooke, "Symmetric functions of a two-way array," *Ann. Math. Stat.*, this issue.

[4] J. W. Tukey, "Interaction in a row-by-column design," Statistical Research Group, Princeton University, Memorandum Report 18, July, 1949.

[5] J. W. Tukey, "Some sampling simplified," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 501–519.

ON THE ESTIMATION OF REGRESSION COEFFICIENTS OF A VECTOR-VALUED TIME SERIES WITH A STATIONARY RESIDUAL[1]

By Murray Rosenblatt[2]

University of Chicago

1. Summary. Time series which are realizations of a vector-valued stochastic process of dimension two with a stationary disturbance are considered. Linear estimates of the regression coefficients of the time series are discussed, in particular the least-squares or classical estimate and the Markov estimate. The least-squares estimate is the estimate computed under the assumption that the components of the disturbance are orthogonal processes and orthogonal to each other. It is known that the Markov estimate is in general better than the least-squares estimate. The asymptotic behavior of the covariance matrices of the least-squares estimate and of the Markov estimate is described. Conditions under which the least-squares estimate is as good asymptotically as the Markov estimate are obtained, that is, conditions under which the least-squares estimate is efficient asymptotically in the class of linear unbiased estimates. The analogues of the results described for vector-valued time series of dimension greater than two can be seen to hold.

2. Introduction. The presentation of the results of this paper is carried out for the case of a two-dimensional process because of the greater simplicity and clarity in exposition. The general n-dimensional case is briefly discussed in Section 9. Let us consider *a two-dimensional complex-valued discrete parameter process*, that is, *a sequence of stochastic vectors*

$$y_t = \begin{pmatrix} {}_1y_t \\ {}_2y_t \end{pmatrix} = x_t + m_t = \begin{pmatrix} {}_1x_t \\ {}_2x_t \end{pmatrix} + \begin{pmatrix} {}_1m_t \\ {}_2m_t \end{pmatrix}, \quad t = \cdots, -1, 0, 1, \cdots, \tag{2.1}$$

where $m_t = Ey_t$ is the mean value sequence and $x_t = y_t - m_t$ is the residual process. We *introduce the covariance sequence* (*x_t' denotes the conjugated transpose of x_t*)

$$E(y_s - m_s)(y_t - m_t)' = Ex_sx_t' = \begin{pmatrix} E\,{}_1x_s\,{}_1\bar{x}_t & E\,{}_1x_s\,{}_2\bar{x}_t \\ E\,{}_2x_s\,{}_1\bar{x}_t & E\,{}_2x_s\,{}_2\bar{x}_t \end{pmatrix} = \begin{pmatrix} {}_{11}r_{s,t} & {}_{12}r_{s,t} \\ {}_{21}r_{s,t} & {}_{22}r_{s,t} \end{pmatrix} = r_{s,t}. \tag{2.2}$$

The assumption that the random variables are complex-valued is made for mathematical convenience. The real-valued case is, of course, the one of greatest

Received October 4, 1954.

[1] Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research.

[2] Now at Indiana University.

statistical interest and is discussed in Section 7 in some detail. Sections 2 and 3 are an extended discussion of the assumptions made in the paper and their motivation. *All the assumptions made in Sections* 2 *and* 3 (except possibly for that of a real-valued time series) *will be held to in all sections except Section* 8. *The residual process x_t is said to be stationary in the wide sense if $r_{s,t} = r_{s-t}$, and I shall assume that this is the case.* Then the covariance sequence has the representation

$$r_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dF(\lambda), \tag{2.3}$$

where $F(\lambda)$ is a matrix-valued function

$$F(\lambda) = \begin{pmatrix} F_{11}(\lambda) & F_{12}(\lambda) \\ F_{21}(\lambda) & F_{22}(\lambda) \end{pmatrix} \tag{2.4}$$

that is nondecreasing; that is, $\Delta F(\lambda) \geqq 0$ (cf. [2]). *The functions $F_{11}(\lambda)$, $F_{22}(\lambda)$ are the spectral distribution functions of ${}_1x_t$, ${}_2x_t$, respectively, while $F_{12}(\lambda)$, $F_{21}(\lambda)$ are the cross-spectral distribution functions of the two coordinates of x_t. We assume that the spectrum is absolutely continuous; that is, that*

$$F_{ij}(\lambda) = \int_{-\pi}^{\lambda} f_{ij}(u)\, du, \qquad i, j = 1, 2, \tag{2.5}$$

and that the spectral densities $f_{ij}(\lambda)$ are continuous. The spectral densities $f_{ii}(\lambda)$, $i = 1, 2$, are assumed to be positive. Note that $f_{12}(\lambda) = \overline{f_{21}(\lambda)}$. The inequality

$$|f_{12}(\lambda)|^2 \leqq f_{11}(\lambda) f_{22}(\lambda) \tag{2.6}$$

obviously holds. *We shall assume that*

$$|f_{12}(\lambda)|^2 < f_{11}(\lambda) f_{22}(\lambda) \tag{2.7}$$

for all λ.

We shall refer to the set of spectra satisfying this set of conditions as the admissible set of spectra. The equality $|f_{12}(\lambda)|^2 \equiv f_{11}(\lambda) f_{22}(\lambda)$ for all λ amounts to a linear relationship between the two coordinates ${}_1x_t$, ${}_2x_t$ of the form ${}_1x_t = \sum_j c_j\, {}_2x_{t-j}$. If the processes ${}_ix_t$ are orthogonal processes, the spectral densities $f_{ii}(\lambda) \equiv \sigma_i^2/2\pi$, $i = 1, 2$. Such processes are sometimes referred to as "white noise." If the processes ${}_1x_t$, ${}_2x_t$ are orthogonal to each other, the cross-spectral density $f_{12}(\lambda) \equiv 0$.

In Section 7 *we shall assume that the process x_t is a real process. This condition imposes additional restraints on the spectrum, specifically that*

$$f_{ii}(\lambda) = f_{ii}(-\lambda), \qquad i = 1, 2, \tag{2.8}$$

and $f_{12}(\lambda) = f_{21}(-\lambda)$. If the process is real, the admissible class of spectra must satisfy these additional restraints.

Let the regression ${}_im_t$, $i = 1, 2$, be of the form

$${}_im_t = \sum_{\nu=1}^{p_i} {}_ic_\nu\, {}_i\varphi_t^{(\nu)}. \tag{2.9}$$

The problem posed is that of estimating the regression coefficients ${}_ic_\nu$ from a time series $y_1, \cdots, y_N$. The regression vectors

$$ {}_i\varphi^{(\nu)} = \begin{pmatrix} {}_i\varphi_1^{(\nu)} \\ \vdots \\ {}_i\varphi_N^{(\nu)} \end{pmatrix} $$

are assumed known. We are interested in unbiased estimates that are linear in the observations ${}_iy_t$, $i = 1, 2$; $t = 1, \cdots, N$. The two linear estimates that we are specifically interested in are the least-squares estimate and the Markov estimate. Let

$$ {}_im = \begin{pmatrix} {}_im_1 \\ \vdots \\ {}_im_N \end{pmatrix}, \qquad {}_iy = \begin{pmatrix} {}_iy_1 \\ \vdots \\ {}_iy_N \end{pmatrix}, \qquad i = 1, 2, $$

and

$$ m = \begin{pmatrix} {}_1m \\ {}_2m \end{pmatrix}, \qquad y = \begin{pmatrix} {}_1y \\ {}_2y \end{pmatrix}. $$

Define the vectors ${}_ic$ and c by

$$ {}_ic = \begin{pmatrix} {}_ic_1 \\ \vdots \\ {}_ic_{p_i} \end{pmatrix}, \qquad i = 1, 2; $$

$$ c = \begin{pmatrix} {}_1c \\ \\ {}_2c \end{pmatrix}. $$

Also define the matrices

$$ {}_i\Phi = ({}_i\varphi^{(1)}, \cdots, {}_i\varphi^{(p_i)}), $$

$$ \Phi = \begin{pmatrix} {}_1\Phi & 0 \\ 0 & {}_2\Phi \end{pmatrix}. $$

The fact that m_t is the mean value of y_t can be written in vector form as

$$ m = Ey = \Phi c. \tag{2.10} $$

The least-squares estimate c_L^* is the estimate that minimizes the quadratic form

$$ (y - m)'(y - m) = (y - \Phi c)'(y - \Phi c); $$

that is, $c_L^* = (\Phi'\Phi)^{-1}\Phi'y$. Note that we are assuming that $\Phi'\Phi$ is nonsingular. The estimate c_L^* is unbiased

$$ Ec_L^* = (\Phi'\Phi)^{-1}\Phi'Ey = (\Phi'\Phi)^{-1}\Phi'\Phi c = c \tag{2.11} $$

and the covariance matrix of c_L^* is

$$ E(c_L^* - c)(c_L^* - c)' = (\Phi'\Phi)^{-1}\Phi'R\Phi(\Phi'\Phi)^{-1}. $$

The matrix R is the covariance matrix of the vector y. Our assumptions concerning the spectrum of the process x_t imply that the matrix R is nonsingular. *The Markov estimate*

$$c_M^* = (\Phi' R^{-1} \Phi)^{-1} \Phi' R^{-1} y.$$

It is also unbiased and its covariance matrix

$$E(c_M^* - c)(c_M^* - c)' = (\Phi' R^{-1} \Phi)^{-1}. \tag{2.12}$$

The Markov estimate is minimum variance among all linear unbiased estimates in the following sense. Consider any unbiased linear estimate $c^* = My$, $Ec^* = M\Phi c = c$; that is, $M\Phi = I$. Its covariance matrix

$$E(c^* - c)(c^* - c)' = MRM'.$$

One can then show that

$$MRM' \geqq (\Phi' R^{-1} \Phi)^{-1}.$$

These remarks about the least-squares and Markov estimates are well known.

We shall investigate the asymptotic behavior of the covariance matrices of the least-squares and the Markov estimate as $N \to \infty$. Note that the least-squares estimate is identical with the Markov estimate when the processes ${}_i x_t$ are orthogonal processes and are orthogonal to each other. It is of considerable interest to find out when the least-squares estimate is asymptotically as good as the Markov estimate, that is, when it is asymptotically efficient in the set of linear unbiased estimates. Whenever we use the phrase asymptotic efficiency we mean asymptotic efficiency in the class of linear unbiased estimates. The least-squares estimate is much easier to compute than the Markov estimate, since it does not require knowledge of the structure of the process x_t. Even if the structure of the process x_t is known, the computation of the inverse R^{-1} may be very tedious. We will discuss the question of asymptotic efficiency. These problems are discussed in [3], [4], and [5] for one-dimensional time series. New aspects of these problems arise in the multidimensional case that we discuss in this paper. The principal results of the paper are given in Sections 4, 5, 6, and 7.

The discussion is based on what might be called a generalized harmonic analysis of the regression vectors. In carrying out this analysis we will have to impose some conditions on the asymptotic behavior of these vectors. However, these conditions will be sufficiently broad to allow most of the usual types of regression sequences. The techniques used are similar to those employed in [5].

3. The regression spectrum. *Let* ${}_i\Phi_N^{(\nu)} = \sum_{t=1}^N | {}_i\varphi_t^{(\nu)} |^2$, $i = 1, 2$. *We first assume that* ${}_i\Phi_N^{(\nu)} \to \infty$ *as* $N \to \infty$. Some condition of this type is required if we are to be able to estimate c consistently. *We also require that*

$$\lim_{N \to \infty} {}_i\Phi_{N+h}^{(\nu)} / {}_i\Phi_N^{(\nu)} = 1 \tag{3.1}$$

for every fixed h. Let the limits

$$(3.2)\qquad {}_{ij}M_h^{(\nu,\mu)} = \lim_{n\to\infty} \sum_{t=1}^{N} \frac{\overline{{}_i\varphi_{t+h}^{(\nu)}}\,{}_j\varphi_t^{(\mu)}}{\sqrt{{}_i\Phi_N^{(\nu)}\,{}_j\Phi_N^{(\mu)}}},$$

$i, j = 1, 2$; $\nu = 1, \cdots, p_i$; $\mu = 1, \cdots, p_j$, *exist for all* $h \geqq 0$. If we set ${}_i\varphi_t^{(\nu)} = 0$ for $t < 0$, it can be seen that the limits ${}_{ij}M_{-h}^{(\nu,\mu)}$, $h > 0$, exist and that

$$(3.3)\qquad {}_{ij}M_h^{(\nu,\mu)} = \overline{{}_{ji}M_h^{(\mu,\nu)}}.$$

Let the matrices

$$(3.4)\qquad {}_{ij}M_h = \{{}_{ij}M_h^{(\nu,\mu)};\ \nu = 1, \cdots, p_i,\ \mu = 1, \cdots, p_j\}, \quad i, j = 1, 2,$$

and

$$M_h = \begin{pmatrix} {}_{11}M_h & {}_{12}M_h \\ {}_{21}M_h & {}_{22}M_h \end{pmatrix}.$$

The matrices M_h, $h = \cdots, -1, 0, 1, \cdots$, form a positive definite sequence; that is, given any $p_1 + p_2$ vector z and any finite vector a

$$\sum_{\nu,\mu} \bar{a}_\nu z' M_{\nu-\mu} z a_\mu \geqq 0.$$

The matrix sequence M_h *then has the representation*

$$(3.5)\qquad M_h = \int_{-\pi}^{\pi} e^{ih\lambda}\, dM(\lambda), \qquad h = \cdots, -1, 0, 1, \cdots,$$

where $M(\lambda)$ *is a matrix-valued function that is nondecreasing so that* $\Delta M(\lambda) \geqq 0$ *for all* λ. *Note that if all the regression vectors are real, we have* $dM(\lambda) = \overline{dM(-\lambda)}$. It will be convenient at times to write $M(\lambda)$ in the form

$$(3.6)\qquad M(\lambda) = \begin{pmatrix} {}_{11}M(\lambda) & {}_{12}M(\lambda) \\ {}_{21}M(\lambda) & {}_{22}M(\lambda) \end{pmatrix},$$

where ${}_{ij}M(\lambda)$ is a $p_i \times p_j$ matrix. It is clear that

$$(3.7)\qquad {}_{ij}M_h = \int_{-\pi}^{\pi} e^{ih\lambda}\, d\,{}_{ij}M(\lambda), \qquad i, j = 1, 2.$$

The matrix-valued functions ${}_{11}M(\lambda)$, ${}_{22}M(\lambda)$ *are nondecreasing and* ${}_{12}M(\lambda) = {}_{21}M(\lambda)'$. *Let*

$$(3.8)\qquad M_0 = M(\pi) - M(-\pi) = M$$

and

$$(3.9)\qquad {}_{ii}M_0 = {}_{ii}M(\pi) - {}_{ii}M(-\pi) = {}_iM, \qquad i = 1, 2.$$

We shall assume that ${}_1M$ *and* ${}_2M$ *are nonsingular.* This means that there is no vector

$$ {}_ia = \begin{pmatrix} {}_ia_1 \\ \vdots \\ {}_ia_{p_i} \end{pmatrix} \neq 0 $$

such that

$$ \sum_{\nu=1}^{p_i} {}_ia_\nu \frac{{}_i\varphi^{(\nu)}}{{}_i\Phi_N^{(\nu)}} \to 0 $$

as $N \to \infty$, $i = 1, 2$. Thus, vectors ${}_i\varphi^{(\nu)}$ are asymptotically linearly independent in the sense described above. Such a condition is required if we are to be able to estimate the regression coefficients consistently as $N \to \infty$. The conditions imposed on the regression vectors are sufficiently broad to include the case of polynomial or trigonometric regression or mixed polynomial and trigonometric regression.

The singular case in which the regression of one component ${}_2m_t \equiv 0$ and one wishes to estimate the regression coefficients ${}_1c_\nu$ of the regression of the other component does not satisfy the conditions imposed on the regression vectors above. It is, however, of some interest and we shall discuss it in Section 8.

Let the diagonal matrix

$$ D_N = \begin{bmatrix} \sqrt{{}_1\Phi_N^{(1)}} & & & & & & 0 \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \sqrt{{}_1\Phi_N^{(p_1)}} & & & \\ & & & & \sqrt{{}_2\Phi_N^{(1)}} & & \\ & & & & & \cdot & \\ 0 & & & & & & \sqrt{{}_2\Phi_N^{(p_2)}} \end{bmatrix} = \begin{pmatrix} {}_1D_N & 0 \\ 0 & {}_2D_N \end{pmatrix}. $$

We shall show that the limits of

$$ D_N E(c_L^* - c)(c_L^* - c)' D_N = D_N(\Phi'\Phi)^{-1} D_N D_N^{-1} \Phi' R \Phi D_N^{-1} D_N (\Phi'\Phi)^{-1} D_N $$

and

$$ D_N E(c_M^* - c)(c_M^* - c)' D_N = D_N(\Phi' R^{-1} \Phi)^{-1} D_N $$

exist as $N \to \infty$ and shall obtain expressions for these limits in terms of the

spectrum of x_t and the regression spectrum (see Sections 4 and 5). It will be convenient for us to write

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}, \tag{3.10}$$

where R_{11}, R_{22} *are the covariance matrices of* ${}_1y$, ${}_2y$ *respectively, while* $R_{12} = R_{21}'$ *is the cross-covariance matrix of* ${}_1y$ *with* ${}_2y$.

Note that the assumptions concerning the regression vectors imply that

$$\lim_{N\to\infty} D_N^{-1}(\Phi'\Phi)D_N^{-1} = \begin{pmatrix} {}_1M & 0 \\ 0 & {}_2M \end{pmatrix}, \tag{3.11}$$

which is nonsingular.

4. The least-squares estimate.

THEOREM 1. *Under the assumptions made in Section 2 on the spectrum* $f(\lambda)$ *of the process* x_t *and the assumptions made in Section 3 on the regression of the process* y_t,

$$\lim_{N\to\infty} D_N E(c_L^* - c)(c_L^* - c)' D_N$$

$$= 2\pi \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix} \begin{bmatrix} \int_{-\pi}^{\pi} f_{11}(-\lambda)\, d\,{}_{11}M(\lambda) & \int_{-\pi}^{\pi} f_{12}(-\lambda)\, d\,{}_{12}M(\lambda) \\ \int_{-\pi}^{\pi} f_{21}(-\lambda)\, d\,{}_{21}M(\lambda) & \int_{-\pi}^{\pi} f_{22}(-\lambda)\, d\,{}_{22}M(\lambda) \end{bmatrix} \times \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix}. \tag{4.1}$$

In discussing the asymptotic behavior of the covariance matrix of the least-squares estimate, it will clearly be enough to consider $D_N^{-1}\Phi'R\Phi D_N^{-1}$. We will approximate R above and below by positive definite matrices of a simpler form.

Consider the quadratic form

$$z'Rz = {}_1z'R_{11}z_1 + {}_2z'R_{21}z_1 + {}_1z'R_{12}z_2 + z_2'R_{22}z,$$

where $z = \begin{pmatrix} {}_1z \\ {}_2z \end{pmatrix}$ and ${}_1z$, ${}_2z$ are N-vectors, so as to see how to approximate R conveniently. Clearly

$$z'Rz = \int_{-\pi}^{\pi} |{}_1z(-\lambda)|^2 f_{11}(\lambda)\, d\lambda + \int_{-\pi}^{\pi} \overline{{}_2z(-\lambda)}\, f_{21}(\lambda)\, {}_1z(-\lambda)\, d\lambda$$
$$+ \int_{-\pi}^{\pi} \overline{{}_1z(-\lambda)}\, f_{12}(\lambda)\, {}_2z(-\lambda)\, d\lambda + \int_{-\pi}^{\pi} |{}_2z(-\lambda)|^2 f_{22}(\lambda)\, d\lambda,$$

where ${}_iz(\lambda) = \sum_{k=1}^{N} {}_iz_k e^{ik\lambda}$, $i = 1, 2$. Note that $z'Rz$ can be written in the more convenient form

$$z'Rz = \int_{-\pi}^{\pi} z(-\lambda)' f(\lambda)\, z(-\lambda)\, d\lambda,$$

where

$$z(\lambda) = \begin{pmatrix} {}_1z(\lambda) \\ {}_2z(\lambda) \end{pmatrix}, \qquad f(\lambda) = \begin{pmatrix} f_{11}(\lambda) & f_{12}(\lambda) \\ f_{21}(\lambda) & f_{22}(\lambda) \end{pmatrix}.$$

Now $|f(\lambda)|$ is nonsingular for all λ, since $f_{11}(\lambda)f_{22}(\lambda) > |f_{12}(\lambda)|^2$ for all λ

Let

$$\begin{pmatrix} a_1 & c_1 \\ \bar{c}_1 & b_1 \end{pmatrix}, \qquad \begin{pmatrix} a_2 & c_2 \\ \bar{c}_2 & b_2 \end{pmatrix}$$

be two positive definite 2×2 matrices. They are positive definite if and only if $a_i, b_i \geqq 0$ and $a_i b_i \geqq |c_i|^2$, $i = 1, 2$. Moreover,

$$\begin{pmatrix} a_1 & c_1 \\ \bar{c}_1 & b_1 \end{pmatrix} \geqq \begin{pmatrix} a_2 & c_2 \\ \bar{c}_2 & b_2 \end{pmatrix}$$

if and only if $a_1 \geqq a_2$, $b_1 \geqq b_2$, and $(a_1 - a_2)(b_1 - b_2) - |c_1 - c_2|^2 \geqq 0$. Now $f_{11}(\lambda), f_{22}(\lambda)$ are positive continuous functions and inequality (2.7) holds. Given any $\epsilon > 0$, we can find finite trigonometric polynomials

$$g_{ij}(\lambda) = \sum_{k=-q}^{q} {}_{ij}g_k\, e^{ik\lambda}, \tag{4.2}$$

$$h_{ij}(\lambda) = \sum_{k=-q}^{q} {}_{ij}h_k\, e^{ik\lambda}, \tag{4.3}$$

$i, j = 1, 2$, satisfying the inequalities

$$\begin{gathered} g_{11}(\lambda) \geqq f_{11}(\lambda) \geqq h_{11}(\lambda) > 0, \\ g_{22}(\lambda) \geqq f_{22}(\lambda) \geqq h_{22}(\lambda) > 0, \\ g_{11}(\lambda)g_{22}(\lambda) > |g_{12}(\lambda)|^2, \\ h_{11}(\lambda)h_{22}(\lambda) > |h_{12}(\lambda)|^2, \\ \epsilon > (g_{11}(\lambda) - f_{11}(\lambda))(g_{22}(\lambda) - f_{22}(\lambda)) > |g_{12}(\lambda) - f_{12}(\lambda)|^2, \\ \epsilon > (f_{11}(\lambda) - h_{11}(\lambda))(f_{22}(\lambda) - h_{22}(\lambda)) > |f_{12}(\lambda) - h_{12}(\lambda)|^2, \\ |g_{ii}(\lambda) - h_{ii}(\lambda)| < \epsilon, \end{gathered}$$

$i = 1, 2$, for all λ. Let

$$\begin{aligned} G_{ij} &= \{{}_{ij}g_{k-l}\,;\, k, l = 1, \cdots, N\}, \\ H_{ij} &= \{{}_{ij}h_{k-l}\,;\, k, l = 1, \cdots, N\}, \end{aligned} \qquad i, j = 1, 2,$$

and

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \qquad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}.$$

Let

$$g(\lambda) = \begin{pmatrix} g_{11}(\lambda) & g_{12}(\lambda) \\ g_{21}(\lambda) & g_{22}(\lambda) \end{pmatrix}, \qquad h(\lambda) = \begin{pmatrix} h_{11}(\lambda) & h_{12}(\lambda) \\ h_{21}(\lambda) & h_{22}(\lambda) \end{pmatrix}.$$

Now

$$\begin{aligned} z'\,Gz &= \int_{-\pi}^{\pi} z(-\lambda)'\,g(\lambda)\,z(-\lambda)\,d\lambda \geqq z'\,Rz \\ &= \int_{-\pi}^{\pi} z(-\lambda)'\,f(\lambda)\,z(-\lambda)\,d\lambda \geqq z'\,Hz \\ &= \int_{-\pi}^{\pi} z(-\lambda)'\,h(\lambda)\,z(-\lambda)\,d\lambda, \end{aligned}$$

so that

$$G \geqq R \geqq H. \tag{4.4}$$

Clearly, $D_N^{-1}\Phi' G\Phi D_N^{-1} \geqq D_N^{-1}\Phi' R\Phi D_N^{-1} \geqq D_N^{-1}\Phi' H\Phi D_N^{-1}$. We shall obtain the limit of $D_N^{-1}\Phi' G\Phi D_N^{-1}$ as $N \to \infty$. This matrix is easier to deal with, since ${}_{ij}g_{k-l} = 0$ if $|k - l| > q$. A typical element of the matrix in question is

$$\sum_{t,\tau=1}^{N} \frac{\overline{{}_i\varphi_t^{(\nu)}}\;{}_{ij}g_{t-\tau}\;{}_j\varphi_\tau^{(\mu)}}{\sqrt{{}_i\Phi_N^{(\nu)}\;{}_j\Phi_N^{(\mu)}}} = \sum_{k=0}^{q} {}_{ij}g_k \sum_{\tau=1}^{N-k} \frac{\overline{{}_i\varphi_{\tau+k}^{(\nu)}}\;{}_j\varphi_\tau^{(\mu)}}{\sqrt{{}_i\Phi_N^{(\nu)}\;{}_j\Phi_N^{(\mu)}}} + \sum_{k=-q}^{-1} {}_{ij}g_k \sum_{\tau=1-k}^{N} \frac{\overline{{}_i\varphi_{\tau+k}^{(\nu)}}\;{}_j\varphi_\tau^{(\mu)}}{\sqrt{{}_i\Phi_N^{(\nu)}\;{}_j\Phi_N^{(\mu)}}},$$

which approaches

$$\sum_k {}_{ij}g_k\;{}_{ij}M_k^{(\nu,\mu)} = 2\pi \int_{-\pi}^{\pi} g_{ij}(-\lambda)\,d\;{}_{ij}M^{(\nu,\mu)}(\lambda)$$

as $N \to \infty$, so that

$$\lim_{N\to\infty} D_N^{-1}\Phi'\,G\Phi\,D_N^{-1} = 2\pi \begin{bmatrix} \int_{-\pi}^{\pi} g_{11}(-\lambda)\,d\;{}_{11}M(\lambda) & \int_{-\pi}^{\pi} g_{12}(-\lambda)\,d\;{}_{12}M(\lambda) \\ \int_{-\pi}^{\pi} g_{21}(-\lambda)\,d\;{}_{21}M(\lambda) & \int_{-\pi}^{\pi} g_{22}(-\lambda)\,d\;{}_{22}M(\lambda) \end{bmatrix}.$$

In like manner, one can show that

$$\lim_{N\to\infty} D_N^{-1}\Phi'\,H\Phi\,D_N^{-1} = 2\pi \begin{bmatrix} \int_{-\pi}^{\pi} h_{11}(-\lambda)\,d\;{}_{11}M(\lambda) & \int_{-\pi}^{\pi} h_{12}(-\lambda)\,d\;{}_{12}M(\lambda) \\ \int_{-\pi}^{\pi} h_{21}(-\lambda)\,d\;{}_{21}M(\lambda) & \int_{-\pi}^{\pi} h_{22}(-\lambda)\,d\;{}_{22}M(\lambda) \end{bmatrix}.$$

Making use of the inequality (4.4), on letting $\epsilon \to 0$ we see that

$$\lim_{N\to\infty} D_N^{-1}\Phi'\,R\Phi D_N^{-1} = 2\pi \begin{bmatrix} \int_{-\pi}^{\pi} f_{11}(-\lambda)\,d\;{}_{11}M(\lambda) & \int_{-\pi}^{\pi} f_{12}(-\lambda)\,d\;{}_{12}M(\lambda) \\ \int_{-\pi}^{\pi} f_{21}(-\lambda)\,d\;{}_{21}M(\lambda) & \int_{-\pi}^{\pi} f_{22}(-\lambda)\,d\;{}_{22}M(\lambda) \end{bmatrix}. \tag{4.5}$$

This limiting matrix will be shown to be nonsingular in Section 5. *Thus* (4.1) is valid.

5. The Markov estimate.

THEOREM 2. Under the assumptions made in Section 2 on the spectrum $f(\lambda)$ of the process x_t and the assumptions made in Section 3 on the regression of y_t,

$$(5.1)\quad \lim_{N\to\infty} D_N E(c_M^* - c)(c_M^* - c)' D_N$$

$$= 2\pi \begin{bmatrix} \int_{-\pi}^{\pi} \frac{f_{22}(-\lambda)\, d\,{}_{11}M(\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2} & -\int_{-\pi}^{\pi} \frac{f_{12}(-\lambda)\, d\,{}_{12}M(\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2} \\ -\int_{-\pi}^{\pi} \frac{f_{21}(-\lambda)\, d\,{}_{21}M(\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2} & \int_{-\pi}^{\pi} \frac{f_{11}(-\lambda)\, d\,{}_{22}M(\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2} \end{bmatrix}^{-1}.$$

In discussing the Markov estimate it will be enough to consider $D_N^{-1}\Phi' R^{-1}\Phi D_N^{-1}$. We shall again approximate R above and below by positive definite matrices of a simpler form. Here we will approximate $f_{11}(\lambda), f_{22}(\lambda)$ by the absolute square of reciprocals of finite trigonometric polynomials, while $f_{12}(\lambda)$ will be approximated as before by a finite trigonometric polynomial. Given any $\epsilon > 0$, we can find finite trigonometric polynomials

$$(5.2)\quad \begin{aligned} \alpha_i(\lambda) &= \sum_{k=-q}^{q} {}_i\alpha_k\, e^{-ik\lambda}, \\ \beta_i(\lambda) &= \sum_{k=-q}^{q} {}_i\beta_k\, e^{-ik\lambda}, \\ \gamma_i(\lambda) &= \sum_{k=-q}^{q} {}_i\gamma_k\, e^{-ik\lambda}, \end{aligned}$$

$i = 1, 2$, satisfying the inequalities

$$(5.3)\quad \begin{gathered} |\alpha_1(\lambda)|^{-2} \geqq f_{11}(\lambda) \geqq |\alpha_2(\lambda)|^{-2}, \\ |\beta_1(\lambda)|^{-2} \geqq f_{22}(\lambda) \geqq |\beta_2(\lambda)|^{-2}, \\ |\alpha_1(\lambda)\beta_1(\lambda)|^{-2} > |\gamma_1(\lambda)|^2, \\ |\alpha_2(\lambda)\beta_2(\lambda)|^{-2} > |\gamma_2(\lambda)|^2, \\ \epsilon > (|\alpha_1(\lambda)|^{-2} - f_{11}(\lambda))(|\beta_1(\lambda)|^{-2} - f_{22}(\lambda)) > |\gamma_1(\lambda) - f_{12}(\lambda)|^2, \\ \epsilon > (f_{11}(\lambda) - |\alpha_2(\lambda)|^{-2})(f_{22}(\lambda) - |\beta_2(\lambda)|^{-2}) > |\gamma_2(\lambda) - f_{12}(\lambda)|^2, \\ |\alpha_1(\lambda)|^{-2} - |\alpha_2(\lambda)|^{-2} < \epsilon, \\ |\beta_1(\lambda)|^{-2} - |\beta_2(\lambda)|^{-2} < \epsilon \end{gathered}$$

for all λ. Let

$${}_iR = \begin{pmatrix} {}_iR_{11} & {}_iR_{12} \\ {}_iR_{21} & {}_iR_{22} \end{pmatrix}$$

be the covariance matrix of y when the process x_t is such that $|\alpha_i(\lambda)|^{-2}$, $|\beta_i(\lambda)|^{-2}$ are the spectral densities of the components ${}_1x_t$, ${}_2x_t$, respectively while $\gamma_i(\lambda)$ is

the cross-spectral density of ${}_1x_t$ and ${}_2x_t$, $i = 1, 2$. It is clear that ${}_1R$ and ${}_2R$ are nonsingular and that

$$ {}_1R \geqq R \geqq {}_2R. \tag{5.4}$$

For the moment let us assume that R is of the same form as one of the matrices ${}_iR$. Then

$$R_{11} = \Delta^{-1}\Delta'^{-1}, \qquad R_{22} = w^{-1}w'^{-1},$$

where $\Delta_{ij} = \alpha_{i-j}$ and $w_{ij} = \beta_{i-j}$ unless $i, j \leqq q$ or

$$N - i, N - j \leqq q \ (\alpha_k = 0, \beta_k = 0, \gamma_k = 0 \text{ if } |k| > q).$$

It is also clear that the (i, j)th element of R_{12} is γ_{i-j}. Let $P = \Delta R_{12} w'$. Then

$$R = \begin{pmatrix} \Delta^{-1}\Delta'^{-1} & \Delta^{-1}Pw'^{-1} \\ w^{-1}P'\Delta'^{-1} & w^{-1}w'^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \Delta^{-1} & 0 \\ 0 & w^{-1} \end{pmatrix}\begin{pmatrix} I & P \\ P' & I \end{pmatrix}\begin{pmatrix} \Delta'^{-1} & 0 \\ 0 & w'^{-1} \end{pmatrix}.$$

It is clear that both

$$\begin{pmatrix} I & P \\ P' & I \end{pmatrix} \text{ and } \begin{pmatrix} I & -P \\ -P' & I \end{pmatrix}$$

are nonnegative definite. Moreover, they commute. Since the matrices commute, it is clear that their product is nonnegative definite and that $I - PP'$, $I - P'P \geqq 0$.

We would like to show that the maximal eigenvalue λ_m of either PP' or $P'P$ is less than one and bounded away from one as $N \to \infty$; that is, $\lambda_m < 1 - \epsilon$ as $N \to \infty$ for some $\epsilon > 0$. This can be seen by noting that the minimal eigenvalues of both

$$\begin{pmatrix} I & P \\ P' & I \end{pmatrix} \text{ and } \begin{pmatrix} I & -P \\ -P' & I \end{pmatrix}$$

are bounded away from zero as $N \to \infty$. It will be enough to show this for

$$\begin{pmatrix} I & P \\ P' & I \end{pmatrix}$$

Let u be any $2N$ vector. Then

$$u'\begin{pmatrix} I & P \\ P' & I \end{pmatrix}u = v'\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}v \geqq \epsilon v'v$$

$$= \epsilon u'\begin{pmatrix} R_{11}^{-1} & 0 \\ 0 & R_{22}^{-1} \end{pmatrix}u = \epsilon\epsilon' u'u$$

as $N \to \infty$, where $\epsilon, \epsilon' > 0$ and $v' = u' \begin{pmatrix} \Delta & 0 \\ 0 & w \end{pmatrix}$. Now

$$R^{-1} = \begin{pmatrix} \Delta' & 0 \\ 0 & w' \end{pmatrix} \begin{pmatrix} I & P \\ P' & I \end{pmatrix}^{-1} \begin{pmatrix} \Delta & 0 \\ 0 & w \end{pmatrix}.$$

But

$$\begin{pmatrix} I & P \\ P' & I \end{pmatrix}^{-1} = \begin{pmatrix} (I - PP')^{-1} & -P(I - P'P)^{-1} \\ -P'(I - PP')^{-1} & (I - P'P)^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} I & -P \\ -P' & I \end{pmatrix} \begin{pmatrix} (I - PP')^{-1} & 0 \\ 0 & (I - P'P)^{-1} \end{pmatrix}.$$

We can write

$$\begin{pmatrix} (I - PP')^{-1} & 0 \\ 0 & (I - P'P)^{-1} \end{pmatrix} = \sum_{k=0}^{\infty} \begin{pmatrix} (PP')^k & 0 \\ 0 & (P'P)^k \end{pmatrix}, \tag{5.5}$$

(5.6) since $0 \leqq PP', P'P \leqq (1 - \epsilon)I$. But then

$$D_N^{-1} \Phi' R^{-1} \Phi D_N^{-1}$$

$$= \sum_{k=0}^{\infty} \begin{pmatrix} {}_1D_N^{-1} \, {}_1\Phi' \Delta' (PP')^k \Delta \, {}_1\Phi \, {}_1D_N^{-1} & -{}_1D_N^{-1} \, {}_1\Phi' \Delta' P (P'P)^k w \, {}_2\Phi \, {}_2D_N^{-1} \\ -{}_2D_N^{-1} \, {}_2\Phi' w' P' (PP')^k \Delta \, {}_1\Phi \, {}_1D_N^{-1} & {}_2D_N^{-1} \, {}_2\Phi' w' (P'P)^k w \, {}_2\Phi \, {}_2D_N^{-1} \end{pmatrix}$$

$$= \sum_{k=0}^{\infty} Q_k.$$

Now the (ν, μ)th element of ${}_1D_N^{-1} \, {}_1\Phi' \Delta' P (P'P)^k w \, {}_2\Phi \, {}_2D_N^{-1}$ is

$$\sum_{t,\tau=1}^{N} \frac{\overline{{}_1\varphi_t^{(\nu)}} h_{t,\tau} \, {}_2\varphi_\tau^{(\mu)}}{\sqrt{{}_1\Phi_N^{(\nu)} \, {}_2\Phi_N^{(\mu)}}},$$

where

$$h_{t,\tau} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-\tau)\lambda} \gamma(\lambda) \, |\alpha(\lambda)|^{2k+2} \, |\beta(\lambda)|^{2k+2} \, |\gamma(\lambda)|^{2k} \, d\lambda$$

$$= s_{t-\tau}$$

unless $t, \tau \leqq 12(k + 1)q$ or $N - t, N - \tau \leqq 12(k + 1)q$. Note that $s_j = 0$ if $|j| > 12(k + 1)q$. It is clear that each of the terms

$$\left| \frac{\overline{{}_1\varphi_t^{(\nu)}} h_{t,\tau} \, {}_2\varphi_\tau^{(\mu)}}{\sqrt{{}_1\Phi_N^{(\nu)} \, {}_2\Phi_N^{(\mu)}}} \right|$$

approaches zero as $N \to \infty$, since $|h_{t,\tau}|$ is uniformly bounded in t, τ, N and

$$|{}_1\varphi_t^{(\nu)}|^2/{}_1\Phi_N^{(\nu)} \to 0$$

as $N \to \infty$ for fixed t. But then

$$\lim_{N\to\infty} \sum_{t,\tau=1}^{N} \frac{\overline{{}_1\varphi_t^{(\nu)}} h_{t,\tau}\, {}_2\varphi_\tau^{(\mu)}}{\sqrt{{}_1\Phi_N^{(\nu)}\, {}_2\Phi_N^{(\mu)}}}$$

$$= \sum_{|k| \leq 12(k+1)q} s_k \lim_{N\to\infty} \sum_{\tau=1}^{N} \frac{\overline{{}_1\varphi_{\tau+k}^{(\nu)}}\, {}_2\varphi_\tau^{(\mu)}}{\sqrt{{}_1\Phi_N^{(\nu)}\, {}_2\Phi_N^{(\mu)}}}$$

$$= \sum_k s_k\, {}_{12}M_k^{(\nu,\mu)} = 2\pi \int_{-\pi}^{\pi} \gamma(-\lambda)\, |\alpha(-\lambda)|^{2k+2}\, |\beta(-\lambda)|^{2k+2}\, |\gamma(-\lambda)|^{2k} \times d\,{}_{12}M^{(\nu,\mu)}(\lambda).$$

But then

$$\lim_{N\to\infty} {}_1D_N^{-1}\, {}_1\Phi'\Delta P(P'P)^k w\, {}_2\Phi\, {}_2D_N^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma(-\lambda)\, |\alpha(-\lambda)|^{2k+2}\, |\beta(-\lambda)|^{2k+2}\, |\gamma(-\lambda)|^{2k}\, d\,{}_{12}M(\lambda).$$

Similarly, one can show that

$$\lim_{N\to\infty} {}_1D_N^{-1}\, {}_1\Phi'\Delta'(PP')^k \Delta\, {}_1\Phi\, {}_1D_N^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\alpha(-\lambda)|^{2k+2}\, |\beta(-\lambda)|^{2k}\, |\gamma(-\lambda)|^{2k}\, d\,{}_{11}M(\lambda)$$

and

$$\lim_{N\to\infty} {}_2D_N^{-1}\, {}_2\Phi' w'(P'P)^k w\, {}_2\Phi\, {}_2D_N^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\alpha(-\lambda)|^{2k}\, |\beta(-\lambda)|^{2k+2}\, |\gamma(-\lambda)|^{2k}\, d\,{}_{22}M(\lambda).$$

Making use of (5.6), it then follows that

$$\lim_{N\to\infty} D_N^{-1}\Phi' R^{-1}\Phi D_N^{-1} = \lim_{N\to\infty} \sum_{k=0}^{\infty} Q_k = \sum_{k=0}^{\infty} \lim_{N\to\infty} Q_k \tag{5.7}$$

$$= \frac{1}{2\pi} \begin{bmatrix} \displaystyle\int_{-\pi}^{\pi} \frac{|\alpha(-\lambda)|^2}{1 - |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2\, |\gamma(-\lambda)|^2}\, d\,{}_{11}M(\lambda) & \displaystyle\int_{-\pi}^{\pi} \frac{\gamma(-\lambda)\, |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2}{1 - |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2\, |\gamma(-\lambda)|^2}\, d\,{}_{12}M(\lambda) \\ \displaystyle\int_{-\pi}^{\pi} \frac{\overline{\gamma(-\lambda)}\, |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2}{1 - |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2\, |\gamma(-\lambda)|^2}\, d\,{}_{21}M(\lambda) & \displaystyle\int_{-\pi}^{\pi} \frac{|\beta(-\lambda)|^2}{1 - |\alpha(-\lambda)|^2\, |\beta(-\lambda)|^2\, |\gamma(-\lambda)|^2}\, d\,{}_{22}M(\lambda) \end{bmatrix}.$$

We have thus found $\lim_{N\to\infty} D_N^{-1}\Phi' R^{-1}\Phi D_N^{-1}$ when R is of the same form as one of the matrices ${}_iR$. Let R now be of the general form, and approximate above and below by ${}_1R$ and ${}_2R$, respectively. On letting $\epsilon \to 0$, we see that

$$\lim_{N\to\infty} D_N^{-1}\Phi' R^{-1}\Phi D_N^{-1}$$

$$(5.8)\qquad = \frac{1}{2\pi}\begin{bmatrix} \displaystyle\int_{-\pi}^{\pi} \frac{f_{22}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d_{11}M(\lambda) & \displaystyle -\int_{-\pi}^{\pi} \frac{f_{12}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d_{12}M(\lambda) \\ \displaystyle -\int_{-\pi}^{\pi} \frac{f_{21}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d_{21}M(\lambda) & \displaystyle\int_{-\pi}^{\pi} \frac{f_{11}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d_{22}M(\lambda) \end{bmatrix}.$$

We can show that the matrix (5.8) is nonsingular. Clearly

$$\begin{pmatrix} \Delta_{11}M(\lambda) & \Delta M_{12}(\lambda) \\ \Delta_{21}M(\lambda) & \Delta_{22}M(\lambda) \end{pmatrix} \geqq 0$$

and $f_{11}(\lambda) f_{22}(\lambda) > |f_{12}(\lambda)|^2$ for all λ. Now

$$f_{22}(-\lambda)z_1'\Delta_{11}M(\lambda)z_1 - f_{12}(-\lambda)z_1'\Delta_{12}M(\lambda)z_2$$

$$-f_{21}(-\lambda)z_2'\Delta_{21}M(\lambda)z_1 + f_{11}(-\lambda)z_2'\Delta_{22}M(\lambda)z_2$$

$$\geqq \left(f_{22}(-\lambda) - \frac{|f_{12}(-\lambda)|^2}{f_{11}(-\lambda)}\right) z_1'\Delta_{11}M(\lambda)z_1,\ \left(f_{11}(-\lambda) - \frac{|f_{12}(-\lambda)|^2}{f_{22}(-\lambda)}\right) z_2'\Delta_{22}M(\lambda)z_2,$$

so that

$$\frac{1}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\begin{pmatrix} f_{22}(-\lambda)\Delta_{11}M(\lambda) & -f_{12}(-\lambda)\Delta_{12}M(\lambda) \\ -f_{21}(-\lambda)\Delta_{21}M(\lambda) & f_{11}(-\lambda)\Delta_{22}M(\lambda) \end{pmatrix}$$

$$\geqq \begin{pmatrix} \Delta_{11}M(\lambda) & 0 \\ 0 & \Delta_{22}M(\lambda) \end{pmatrix},$$

where $\epsilon > 0$. But then matrix (5.8) is greater than or equal to

$$\frac{1}{2\pi}\,\epsilon \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$$

and hence is nonsingular. Relation (5.1) is valid. It follows that the limiting matrix (4.5) is also nonsingular.

6. Asymptotic efficiency of least-squares estimate among linear unbiased estimates when observed process is complex-valued. We want to find out

for what types of regression the least-squares estimate is asymptotically efficient among the linear unbiased estimates for any admissible spectral density matrix $f(\lambda)$. This amounts to asking for the conditions on $M(\lambda)$ such that matrix (4.1) is equal to matrix (5.1); that is,

$$\begin{bmatrix} \int_{-\pi}^{\pi} f_{11}(-\lambda)\, d\,{}_{11}M(\lambda) & \int_{-\pi}^{\pi} f_{12}(-\lambda)\, d\,{}_{12}M(\lambda) \\ \int_{-\pi}^{\pi} f_{21}(-\lambda)\, d\,{}_{21}M(\lambda) & \int_{-\pi}^{\pi} f_{22}(-\lambda)\, d\,{}_{22}M(\lambda) \end{bmatrix} \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix} \times \tag{6.1}$$

$$\begin{bmatrix} \int_{-\pi}^{\pi} \frac{f_{22}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d\,{}_{11}M(\lambda) & -\int_{-\pi}^{\pi} \frac{f_{12}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d\,{}_{12}M(\lambda) \\ -\int_{-\pi}^{\pi} \frac{f_{21}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d\,{}_{21}M(\lambda) & \int_{-\pi}^{\pi} \frac{f_{11}(-\lambda)}{f_{11}(-\lambda)f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d\,{}_{22}M(\lambda) \end{bmatrix} = \begin{pmatrix} {}_1M & 0 \\ 0 & {}_2M \end{pmatrix}.$$

Let us first see what restraints are imposed on the regression spectrum if we require asymptotic efficiency of the least-squares estimate in the smaller class of spectra $f(\lambda)$ where there is no cross-correlation, that is, where $f_{12}(\lambda) \equiv 0$. Since ${}_1y_t$ and ${}_2y_t$ are uncorrelated, they can be treated separately. We make use of the results of [5] where the problem of estimating the regression coefficients of a 1-dimensional process with stationary residuals is discussed. The following restraints on the regression spectrum follow immediately from these results. The nondecreasing function ${}_{ii}M(\lambda)$ increases only on a finite set of points ${}_i\lambda_j$, $j = 1, \cdots, q_i$, where $q_i \leqq p_i$, $i = 1, 2$. The jump of ${}_{ii}M(\lambda)$ at ${}_i\lambda_j$ is $\Delta\,{}_{ii}M({}_i\lambda_j) = {}_{ii}M({}_i\lambda_j+) - {}_{ii}M({}_i\lambda_j-) > 0$ and

$$\Delta\,{}_{ii}M({}_i\lambda_j)\,{}_iM^{-1}\Delta\,{}_{ii}M({}_i\lambda_k) = \delta_{jk}\Delta\,{}_{ii}M({}_i\lambda_j), \qquad i = 1, 2. \tag{6.2}$$

The sum of the ranks of the matrices $\Delta\,{}_{ii}M({}_i\lambda_j)$, $j = 1, \cdots, q_i$, is p_i. It is then clear that the set of points of increase of the nondecreasing function $M(\lambda)$ is the set of points $\{\lambda_j\}$ consisting of the points ${}_1\lambda_k$, $k = 1, \cdots, q_1$, and ${}_2\lambda_k$, $k = 1, \cdots, q_2$. For convenience let ${}_{ij}M_k = {}_{ij}M(\lambda_k+) - {}_{ij}M(\lambda_k-)$. Relations (6.2) can then be rewritten ${}_{ii}M_j\,{}_iM^{-1}{}_{ii}M_k = \delta_{jk}\,{}_{ii}M_j$, $i = 1, 2$. Here either ${}_{11}M_j > 0$ or ${}_{22}M_j > 0$. We shall obtain additional restraints on the regression spectrum and thereby show that the sets of points $\{\lambda_j\}$, $\{{}_1\lambda_j\}$ and $\{{}_2\lambda_j\}$ are the same.

Let us now see what additional restraints on the regression spectrum are implied by asymptotic efficiency of the least-squares estimate when the spec-

trum is such that $f_{11}(\lambda) \equiv f_{22}(\lambda) \equiv 1$ and $f_{12}(-\lambda_j) = \alpha_j$, $|\alpha_j|^2 < 1$. The condition for asymptotic efficiency is then

$$\sum_{j,k} \begin{pmatrix} {}_{11}M_j & \alpha_j\, {}_{12}M_j \\ \bar{\alpha}_j\, {}_{21}M_j & {}_{22}M_j \end{pmatrix} \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix} \frac{1}{1-|\alpha_k|^2} \begin{pmatrix} {}_{11}M_k & -\alpha_k\, {}_{12}M_k \\ -\bar{\alpha}_k\, {}_{21}M_k & {}_{22}M_k \end{pmatrix} = \begin{pmatrix} {}_1M & 0 \\ 0 & {}_2M \end{pmatrix}. \tag{6.3}$$

If all the α's are zero except for one, say α_j, equation (6.3) reduces to

$$\begin{pmatrix} {}_{11}M_j & -\alpha_j\, {}_{12}M_j \\ -\bar{\alpha}_j\, {}_{21}M_j & {}_{22}M_j \end{pmatrix} - \begin{pmatrix} {}_{12}M_j\, {}_2M^{-1}\, {}_{21}M_j & -\alpha_j\, {}_{12}M_j\, {}_2M^{-1}\, {}_{22}M_j \\ -\bar{\alpha}_j\, {}_{21}M_j\, {}_1M^{-1}\, {}_{11}M_j & {}_{21}M_j\, {}_1M^{-1}\, {}_{12}M_j \end{pmatrix} = 0,$$

so that ${}_{ik}M_j\, {}_kM^{-1}\, {}_{kl}M_j = {}_{il}M_j$, $i, k, l = 1, 2$, $i \neq k$. If all the α's are zero except for two, say α_j and α_k, equation (6.3) reduces to

$$\begin{aligned} &\frac{1}{1-|\alpha_k|^2} \begin{pmatrix} 0 & \alpha_j\, {}_{12}M_j \\ \bar{\alpha}_j\, {}_{21}M_j & 0 \end{pmatrix} \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix} \begin{pmatrix} |\alpha_k|^2\, {}_{11}M_k & -\alpha_k\, {}_{12}M_k \\ -\bar{\alpha}_k\, {}_{21}M_k & |\alpha_k|^2\, {}_{22}M_k \end{pmatrix} \\ &+ \frac{1}{1-|\alpha_j|^2} \begin{pmatrix} 0 & \alpha_k\, {}_{12}M_k \\ \bar{\alpha}_k\, {}_{21}M_k & 0 \end{pmatrix} \begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1} \end{pmatrix} \begin{pmatrix} |\alpha_j|^2\, {}_{11}M_j & -\alpha_j\, {}_{12}M_j \\ -\bar{\alpha}_j\, {}_{21}M_j & |\alpha_j|^2\, {}_{22}M_j \end{pmatrix} = 0. \end{aligned} \tag{6.4}$$

But equation (6.4) cannot hold for all values of α_j, α_k less than one in absolute value unless

$${}_{il}M_j\, {}_lM^{-1}\, {}_{ls}M_k = 0, \quad j \neq k; \quad i, l, s = 1, 2; \quad i \neq l.$$

All other relations of this type can be obtained analogously from the matrix equation resulting from the interchange of the first and third matrix on the left side of equation (6.3). This equation obviously also holds if the least-squares estimate is asymptotically efficient. All the restraints on the matrices ${}_{ij}M_k$ can be written briefly

$${}_{ij}M_k\, {}_jM^{-1}\, {}_{jl}M_s = \delta_{ks}\, {}_{il}M_s, \tag{6.5}$$

where $i, j, l = 1, 2$ and $k, s = 1, \cdots, q$ where $q = q_1 = q_2$. It is clear that equations (6.5) cannot hold unless both ${}_{11}M_k$, ${}_{22}M_k > 0$ and have the same rank. Thus asymptotic efficiency of the least-squares estimate implies that $p_1 = p_2$.

THEOREM 3. *The following conditions are necessary if the least-squares estimate is to be asymptotically efficient for all admissible $f(\lambda)$. The function $M(\lambda)$ is a jump unction with a finite number of jumps $\lambda_1, \cdots, \lambda_q$, where $q \leq p = p_1 = p_2$. Let the jumps be*

$${}_{ij}M_k = {}_{ij}M(\lambda_k +) - {}_{ij}M(\lambda_k -), \qquad i, j = 1, 2.$$

Then ($A > 0$ if A is a nonnegative definite matrix but not the null matrix) ${}_{ii}M_k > 0$,[2] $i = 1, 2$ and $k = 1, \cdots, q$ *and*

$${}_{ij}M_k\, {}_jM^{-1}\, {}_{jl}M_s = \delta_{ks}\, {}_{il}M_s, \tag{6.6}$$

where $i, j, l = 1, 2$ and $k, s = 1, \cdots, q$. ${}_{11}M_k$ and ${}_{22}M_k$ have the same rank. The sum of the ranks of ${}_{ii}M_k$, $k = 1, \cdots, q$, is p, $i = 1, 2$. These conditions can easily be seen to be sufficient for the least-squares estimate to be asymptotically efficient for all admissible $f(\lambda)$.

It is of especial interest to consider the case in which both components ${}_1y_t$, ${}_2y_t$ of the observed process have a mixed trigonometric and polynomial regression and the regression vectors of both components are the same; that is,

$$
\begin{aligned}
{}_1\varphi_t^{(\nu)} &= {}_2\varphi_t^{(\nu)} = t^\nu e^{-it\lambda_1}, && \nu = 0, 1, \cdots, s_1, \\
{}_1\varphi_t^{(s_1+1+\nu)} &= {}_2\varphi_t^{(s_1+1+\nu)} = t^\nu e^{-it\lambda_2}, && \nu = 0, 1, \cdots, s_2, \\
{}_1\varphi_t^{(s_1+\cdots+s_{u-1}+u-1+\nu)} &= {}_2\varphi_t^{(s_1+\cdots+s_{u-1}+u-1+\nu)} \\
&= t^\nu e^{-it\lambda_q}, && \nu = 0, 1, \cdots, s_q,
\end{aligned}
$$

where $\lambda_1, \cdots, \lambda_q$ are distinct. *The least-squares estimate can be seen to be asymptotically efficient in the case of such a regression.* The jumps of $M(\lambda)$ are at $\lambda_1, \cdots, \lambda_q$ and

$$
{}_{ij}M_k = \begin{bmatrix} 0 & 0 & 0 \\ 0 & M_k & 0 \\ 0 & 0 & 0 \end{bmatrix},
$$

where

$$
M_k = \left\{ \frac{\sqrt{(2\mu + 1)(2\nu + 1)}}{\mu + \nu + 1} ; \mu, \nu = 0, 1, \cdots, s_k \right\}
$$

and the null submatrix in the upper left-hand corner of ${}_{ij}M_k$ is of order $\sum_{i=1}^{k-1} (s_i + 1)$. It is clear that equations (5.6) are satisfied in this case. *One should note that if the regression vectors of the two components are unequal in number, the least-squares estimate is not asymptotically efficient.* This, for example, would be the case if one component had a linear regression and the other a quadratic regression.

7. Asymptotic efficiency of the least squares estimate when the observed process is real-valued. The case of greatest interest is that in which the process y_t and the regression vectors are real. This condition imposes additional restraints on the spectrum of the process and the regression spectrum. Then

$$
\begin{aligned}
f_{ii}(\lambda) &= f_{ii}(-\lambda), && i = 1, 2, \\
f_{12}(\lambda) &= f_{21}(-\lambda),
\end{aligned}
\tag{7.1}
$$

and

$$
dM(\lambda) = \overline{dM(-\lambda)}. \tag{7.2}
$$

We shall obtain necessary and sufficient conditions on the regression spectrum for the least-squares estimate to be asymptotically efficient for such a process. The derivation of these conditions is analogous to that followed in Section 6.

Just as in Section 6 one can see that asymptotic efficiency of the least-squares estimate implies that there are only a finite number of points of increase of $M(\lambda)$. Because of (7.2) we need only consider the nonnegative points of increase of $M(\lambda)$. Let the nonnegative points of increase be $\lambda_1 = 0 < \lambda_2 < \cdots < \lambda_q$. Of course zero needn't be one of the points of increase but we include it because if it is, the condition on the jump at zero is different from that on jumps at other points. Let ${}_{ij}M_k$ denote the jump of ${}_{ij}M(\lambda)$ at λ_k. Given the matrix A, let $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ be the matrices whose elements are the real and imaginary parts, respectively, of the corresponding elements of A. Equation (6.1) can then be rewritten as

$$\sum_j{}' 2\begin{pmatrix} {}_{11}f_j \operatorname{Re}({}_{11}M_j) & \operatorname{Re}({}_{12}f_j\,{}_{12}M_j) \\ \operatorname{Re}({}_{21}f_j\,{}_{21}M_j) & {}_{22}f_j \operatorname{Re}{}_{22}M_j \end{pmatrix}\begin{pmatrix} {}_1M^{-1} & 0 \\ 0 & {}_2M^{-1}\end{pmatrix} \times$$
$$\sum_k{}' \frac{2}{{}_{11}f_k\,{}_{22}f_k - |{}_{12}f_k|^2}\begin{pmatrix} {}_{22}f_k \operatorname{Re}({}_{11}M_k) & -\operatorname{Re}({}_{12}f_k\,{}_{12}M_k) \\ -\operatorname{Re}({}_{21}f_k\,{}_{21}M_k) & {}_{11}f_k \operatorname{Re}({}_{22}M_k)\end{pmatrix} = \begin{pmatrix} {}_1M & 0 \\ 0 & {}_2M \end{pmatrix}, \tag{7.3}$$

making use of (7.1) and (7.2). Here, ${}_{ij}f_k = {}_{ij}f(\lambda_k)$. The primed summations indicate that the coefficient 2 in the summation is to be replaced by coefficient one when either j or $k = 1$, since $\lambda_1 = 0$. Because of (7.2) we can see that the matrices ${}_{ij}M_1$ have real elements. The fact that $f_{12}(\lambda) = \overline{f_{21}(\lambda)}$ indicates that ${}_{ij}f_1$, $i, j = 1, 2$, is real. Now equation (7.3) is assumed to hold for all ${}_{11}f_j$, ${}_{22}f_j > 0$ and all ${}_{12}f_j$ such that $|{}_{12}f_j|^2 < {}_{11}f_j\,{}_{22}f_j$. A discussion of equation (7.3) analogous to that carried out in Section 6 indicates that the *equation cannot be valid under these conditions unless the following restraints on the matrices* ${}_{ij}M_k$ *are satisfied*:

$$2\operatorname{Re}({}_{ij}M_l)\,{}_jM^{-1}\operatorname{Re}({}_{jk}M_s) = \delta_{ls}\operatorname{Re}({}_{ik}M_l)$$

if $l \neq 0$;

$$\begin{aligned} {}_{ij}M_1\,{}_jM^{-1}\,{}_{jk}M_1 &= {}_{ik}M_1, \\ 2\mathrm{Im}({}_{ij}M_l)\,{}_jM^{-1}\,\mathrm{Im}({}_{ji}M_k) &= -\delta_{lk}\operatorname{Re}({}_{ii}M_l), \end{aligned} \tag{7.4}$$

if $i \neq j$ *and* $l, k \neq 1$, *since* $\mathrm{Im}({}_{ij}M_1) = 0$;

$$\operatorname{Re}({}_{ij}M_k)\,{}_jM^{-1}\,\mathrm{Im}({}_{jl}M_k) = \mathrm{Im}({}_{jl}M_k)\,{}_lM^{-1}\operatorname{Re}({}_{ls}M_k),$$

where $j \neq l$ *and* $k \neq 1$; *and finally,*

$$\operatorname{Re}({}_{ij}M_k)\,{}_jM^{-1}\,\mathrm{Im}({}_{jl}M_s) = \mathrm{Im}({}_{jl}M_k)\,{}_lM^{-1}\operatorname{Re}({}_{lu}M_s) = 0$$

if $k \neq s$. *It can also be readily seen that equation* (7.3) *will be satisfied for all admissible spectra of the process if the conditions* (7.4) *just derived are satisfied.*

THEOREM 4. *The least-squares estimate is asymptotically efficient for all admissible spectra of the process if and only if the regression spectrum is a jump spectrum with a finite number of jumps and the matrices* ${}_{ij}M_k$ *satisfy the conditions* (7.4).

It is again of special interest to consider the case in which both components ${}_1y_t$, ${}_2y_t$ of the observed process have a mixed trigonometric and polynomial regression and the regression vectors of both component are the same, so that

$$
\begin{aligned}
{}_1\varphi_t^{(\nu)} &= {}_2\varphi_t^{(\nu)} = t^\nu, && \nu = 0, 1, \cdots, s_1, \\
{}_1\varphi_t^{(s_1+1+\nu)} &= {}_2\varphi_t^{(s_1+1+\nu)} = t^\nu \cos t\lambda_2, && \nu = 0, 1, \cdots, s_2, \\
{}_1\varphi_t^{(s_1+s_2+2+\nu)} &= {}_2\varphi_k^{(s_1+s_2+2+\nu)} = t^\nu \sin t\lambda_2, && \nu = 0, 1, \cdots, s_2.
\end{aligned}
$$

. . . .

$$
{}_1\varphi_t^{(s_1+2s_2+\cdots+2s_{u-1}+2u-1+\nu)} = {}_2\varphi_t^{(s_1+2s_2+\cdots+2s_{u-1}2u-1+\nu)} = t^\nu \cos t\lambda_u,
$$

$$
\nu = 0, 1, \cdots, s_u,
$$

$$
{}_1\varphi_t^{(s_1+2s_2+\cdots+2s_{u-1}+s_u+2u+\nu)} = {}_2\varphi_t^{(s_1+2s_2+\cdots+2s_{u-1}+s_u+2u+\nu)} = t^\nu \sin t\lambda_u,
$$

$$
\nu = 0, 1, \cdots, s_u,
$$

where $\lambda_1 = 0 < \lambda_2 < \cdots < \lambda_u$. The jumps of $M(\lambda)$ are at 0, $\pm\lambda_2, \cdots, \pm\lambda_u$. We need only discuss the jumps at nonnegative λ, since $dM(\lambda) = \overline{dM(-\lambda)}$. Now

$$
{}_{jk}M_1 = \begin{bmatrix} M_1 & 0 \\ 0 & 0 \end{bmatrix},
$$

where

$$
M_1 = \left\{ \frac{\sqrt{(2\mu+1)(2\nu+1)}}{\mu+\nu+1} ; \mu, \nu = 0, 1, \cdots, s_1 \right\},
$$

$$
{}_{jk}M_l = \tfrac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & M_l & -iM_l & 0 \\ 0 & iM_l & M_l & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},
$$

where

$$
M_l = \left\{ \frac{\sqrt{(2\mu+1)(2\nu+1)}}{\mu+\nu+1} ; \mu, \nu = 0, 1, \cdots, s_l \right\}, \qquad l \neq 1,
$$

and the null submatrix in the upper left-hand corner of ${}_{jk}M_l$ is of order $s_1 + 1 + 2\sum_{h=2}^{l-1}(s_h + 1)$. It is clear that *the least-squares estimate is asymptotically efficient in the case of such a regression, since the conditions* (7.4) *are satisfied.* As in the case of a complex-valued process, one does not have asymptotic efficiency of the least-squares estimate if the regression vectors of the two components are unequal in number. There is, however, an additional restriction that enters into this context and did not arise in Section 6. *If one of the terms in*

the regression is $t^\nu \cos t\lambda$, $\lambda \neq 0$, *one must also have the term* $t^\nu \sin t\lambda$ *for asymptotic efficiency of the least-squares estimate.* Thus, one does not have asymptotic efficiency of the least-squares estimate in the case of the regression

$$ {}_1m_t = {}_1c \cos t\lambda, \qquad {}_2m_t = {}_2c \cos t\lambda, \qquad \lambda \neq 0, $$

and one does in the case of the regression

$$ \begin{aligned} {}_1m_t &= {}_1c_1 \cos t\lambda + {}_1c_2 \sin t\lambda, \\ {}_2m_t &= {}_2c_1 \cos t\lambda + {}_2c_2 \sin t\lambda, \qquad \lambda \neq 0. \end{aligned} $$

8. The Markov and least-squares estimate when the regression of one component vanishes. The special case in which the regression of one component vanishes, say ${}_2m_t \equiv 0$, has not yet been discussed but is of some interest. It is clear that the least-squares estimate can not be asymptotically efficient for all admissible spectra of the observed process in this case. Let c^* now denote a linear unbiased estimate of ${}_1c$. The least-squares estimate of ${}_1c$ in this case is

$$ c_L^* = ({}_1\Phi' {}_1\Phi)^{-1} {}_1\Phi' {}_1y. \tag{8.1} $$

The covariance matrix of the least-squares estimate is

$$ E(c_L^* - {}_1c)(c_L^* - {}_1c)' = ({}_1\Phi' {}_1\Phi)^{-1} {}_1\Phi' R_{11}\, {}_1\Phi ({}_1\Phi' {}_1\Phi)^{-1}. \tag{8.2} $$

The Markov estimate c_M^* of ${}_1c$ is

$$ c_M^* = \left(({}_1\Phi' \; 0) R^{-1} \begin{pmatrix} {}_1\Phi \\ 0 \end{pmatrix} \right)^{-1} ({}_1\Phi' \; 0) R^{-1} y. \tag{8.3} $$

The covariance matrix of the Markov estimate is

$$ \begin{aligned} E(c_M^* - {}_1c)(c_M^* - {}_1c)' &= \left(({}_1\Phi' \; 0) R^{-1} \begin{pmatrix} {}_1\Phi \\ 0 \end{pmatrix} \right)^{-1} \\ &= ({}_1\Phi' (R_{11} - R_{12} R_{22}^{-1} R_{21})^{-1} {}_1\Phi^{-1}). \end{aligned} \tag{8.4} $$

By using techniques analogous to those of Sections 4 and 5, one can show that

$$ \lim_{N \to \infty} {}_1D_N E(c_L^* - {}_1c)(c_L^* - {}_1c)' {}_1D_N = 2\pi\, {}_1M^{-1} \int_{-\pi}^{\pi} f_{11}(-\lambda)\, d\, {}_{11}M(\lambda)\, {}_1M^{-1}, \tag{8.5} $$

while

$$ \begin{aligned} &\lim_{N \to \infty} {}_1D_N E(c_M^* - {}_1c)(c_M^* - c_1)' {}_1D_N \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} \frac{f_{22}(-\lambda)}{f_{11}(-\lambda) f_{22}(-\lambda) - |f_{12}(-\lambda)|^2}\, d\, {}_{11}M(\lambda) \right)^{-1}. \end{aligned} \tag{8.6} $$

Of course c_M^* is the best of all linear unbiased estimates of ${}_1c$ in that it has the

smallest covariance matrix of them all. It is interesting to compare the two estimates when ${}_1m_t \equiv c$, a constant. Then

$${}_1M(\pi) - {}_1M(-\pi) = {}_1M(0+) - {}_1M(0-) = 1,$$

so that

$$\lim_{N\to\infty} \frac{E(c_M^* - c)^2}{E(c^* - c)^2} = 1 - \frac{|f_{12}(0)|^2}{f_{11}(0)f_{22}(0)}.$$

Some aspects of this example have been discussed in [1] from the point of view of discriminant analysis.

9. Processes of dimension higher than two. We shall discuss briefly the case of an n-dimensional process y_t with stationary residuals, $n \geqq 3$, and indicate that most of the results obtained in the two-dimensional case are still valid. Assumptions analogous to those of Sections 2 and 3 are made. Let

$${}_im_t = \sum_{\nu=1}^{p_i} {}_ic_\nu \varphi_t^{(\nu)}, \qquad i = 1, \cdots, n,$$

be the regression of the ith component ${}_iy_t$ of y_t. The spectrum of the residual process $x_t = y_t - m_t$ is again assumed absolutely continuous with continuous spectral densities $f_{ij}(\lambda)$, $i, j = 1, \cdots, n$. Of course $f_{ij}(\lambda)$ is the cross-spectral density of ${}_ix_t$ and ${}_jx_t$. The determinant $|f(\lambda)|$ of the matrix

$$f(\lambda) = \{f_{ij}(\lambda); i, j = 1, \cdots, n\} \tag{9.1}$$

is assumed to be positive for all λ.

Let

$${}_ic = \begin{pmatrix} {}_ic_1 \\ \vdots \\ {}_ic_{p_i} \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} {}_1c \\ \vdots \\ {}_nc \end{pmatrix}.$$

Let c_L^* and c_M^* be the least-squares estimate and the Markov estimate of c, respectively, in terms of the observed process $y_1, \cdots, y_N$. The matrices ${}_i\Phi$, $i = 1, \cdots, n$, are defined just as in Section 2, and we set

$$\Phi = \begin{bmatrix} {}_1\Phi & & 0 \\ & \ddots & \\ 0 & & {}_n\Phi \end{bmatrix}.$$

R is the covariance matrix of the vector $\begin{pmatrix} {}_1y \\ \vdots \\ {}_ny \end{pmatrix}$. The expressions given for c_L^* and c_M^* and their respective covariance matrices in Section 1 are still valid.

The cross-spectral distribution function of the regression vectors of ${}_iy_t$ and ${}_jy_t$ are computed just as in the two-dimensional case. The matrices

${}_{ii}M(\pi) - {}_{ii}M(-\pi) = {}_iM$, $i = 1, \cdots, n$, are assumed to be nonsingular. Let $M(\lambda) = \{{}_{ij}M(\lambda); i, j = 1, \cdots, n\}$. We set

$$D_N = \begin{pmatrix} {}_1D_N & & 0 \\ & \ddots & \\ 0 & & {}_nD_N \end{pmatrix}.$$

Let $\theta_{ij}(\lambda)$ be the cofactor of the element $f_{ij}(\lambda)$ in the matrix $f(\lambda)$. By using an elaboration of the methods of Sections 4 and 5, one can show that

$$\begin{aligned} &\lim_{N\to\infty} D_N E(c_L^* - c)(c_L^* - c)' D_N \\ &= 2\pi M^{-1} \left\{ \int_{-\pi}^{\pi} f_{ij}(-\lambda)\, d_{ij} M(\lambda); i, j = 1, \cdots, n \right\} M^{-1} \end{aligned} \tag{9.2}$$

and

$$\begin{aligned} &\lim_{N\to\infty} D_N E(c_M^* - c)(c_M^* - c)' \\ &= \frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} \theta_{ij}(-\lambda)\, d_{ij} M(\lambda); i, j = 1, \cdots, n \right\}^{-1}. \end{aligned} \tag{9.3}$$

The analogues of the results obtained in Sections 6 and 7 on asymptotic efficiency of the least-squares estimate c_L^* are valid in the n-dimensional case and can be proved in the same way.

10. Concluding remarks. It would be very interesting to find out for what sample size N the various results obtained on asymptotic efficiency of the least-squares estimate are practically valid. An effective way to find out would be to set up a computational program for the calculation of the covariance matrices of the least-squares and Markov estimates for a variety of interesting regressions and spectra $f(\lambda)$. The approximations derived in this paper for these covariance matrices should also be computed and compared with the true covariance matrices.

Consider the case of a real two-dimensional process $y_t = \begin{pmatrix} {}_1y_t \\ {}_2y_t \end{pmatrix}$ where both components have constant mean values $E\,{}_iy_t = {}_ic$, $i = 1, 2$, which we want to estimate. The simplest case of cross-correlation, and a rather uninteresting one, is that in which

$$\operatorname{cov}({}_iy_t, {}_iy_\tau) = \delta_{t\tau},$$

$$\operatorname{cov}({}_1y_t, {}_2y_\tau) = \rho\delta_{t\tau},$$

where $|\rho| < 1$. Nonetheless, it is amusing to note that in this case the least-squares estimate is efficient for all finite N and that

$$\begin{aligned} E(c_L^* - c)(c_L^* - c)' &= E(c_M^* - c)(c_M^* - c)' \\ &= 2\pi D_N^{-1} M^{-1} \left\{ \int_{-\pi}^{\pi} f_{ij}(-\lambda)\, d\, {}_{ij}M(\lambda); i, j = 1, 2 \right\} M^{-1} D_N^{-1} \\ &= 2\pi D_N^{-1} \left\{ \int_{-\pi}^{\pi} \theta_{ij}(-\lambda)\, d\, {}_{ij}M(\lambda); i, j = 1, 2 \right\}^{-1} D_N^{-1} \\ &= \frac{1}{N} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \end{aligned}$$

REFERENCES

[1] W. G. Cochran and C. I. Bliss, "Discriminant functions with covariance," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 151–176.

[2] H. Cramér, "On the theory of stationary random processes," *Ann. Math. Stat.*, Vol. 41 (1940), pp. 215–230.

[3] U. Grenander, "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 252–272.

[4] U. Grenander and M. Rosenblatt, "An extension of a theorem of G. Szegö and its application to the study of stochsatic processes," *Trans. Amer. Math. Soc.*, Vol. 76 (1954), pp. 112–126.

[5] U. Grenander and M. Rosenblatt, "Regression analysis of time series with a stationary disturbance," *Proc. Nat. Acad. Sci. U.S.* (to be published), Vol. 40 (1954), pp. 812–816.

AN APPLICATION OF INFORMATION THEORY TO MULTIVARIATE ANALYSIS, II

By S. Kullback

The George Washington University

0. Summary. Certain results of information theory are applied to some problems of multivariate analysis, including the multivariate linear hypothesis and the hypothesis of homogeneity of covariance matrices. A discussion of certain related linear discriminant functions is also included. Some asymptotic distributions on the null hypothesis are derived. Related problems, still under investigation, are mentioned. The procedures are based on the principle of maximizing information. For the cases considered, the estimates of $I(1:2)$ and $J(1,2)$ turn out to be those obtained by replacing the parameters by unbiased estimates, appropriate to the hypotheses under consideration.

1. Introduction. In a previous paper [20], the author considered certain results of information theory as applied to multivariate normal populations. In particular there was examined the problem of finding the "best" linear function for discriminating between two normal populations, assuming equal means but different population covariance matrices. The multivariate analysis techniques of discriminant analysis, principal components, and canonical correlations were seen to be closely related concepts. (Greenhouse [12], using the information-theory approach, has examined the problem of finding the "best" linear function for discriminating between two multivariate normal populations, with no restrictive assumptions as to means or covariance matrices.)

In [20], the discussion was in terms of population parameters, and questions of estimation and distribution were omitted. In addition to discussing some of the problems of estimation and distribution herein, we also want to consider further application of information theory, and the relation with previous developments, by studying the following four multivariate problems (cf. Roy [28], Section 5.1):

(a) The hypothesis that a k-variate normal population has the covariance matrix σ;

(b) The hypothesis of equality of r means for each of k variates for r k-variate normal populations with different covariance matrices, and with the same covariance matrix;

(c) The multivariate linear hypothesis, including the case of a subhypothesis;

(d) The hypothesis of equality of the covariance matrices of r k-variate normal populations.

The reader is referred to Section 2 of [20] in which the information measures are defined and their properties summarized, with particular reference to prop-

Received March 26, 1954; revised September 8, 1955.

erty (iii) on p. 90 of [20]. (Proofs may be found in [22].) Based on the non-decreasing property of $I(1:2)$ and $J(1,2)$ for sufficient statistics, we follow a principle that may be termed maximizing information in order to attain sufficiency or near sufficiency. It seems intuitively reasonable that such an approach should have certain optimum properties. In certain cases the results are closely related to likelihood ratio tests and this relation is under investigation for the general case. Asymptotic distributions occurring herein verify conclusions derivable from a general asymptotic theory, the details of which are in preparation (cf. Wilks [32]), and in two cases a better approximation than the general theory provides is derived.

It might be remarked that appropriate multivariate extensions of the results in Cramér [6], particularly pp. 11 and 12, and Daniels [7] also provide an alternative basis for a general asymptotic theory.

There are few general (automatic) procedures for finding test criteria. The approach using information theory as a means of determining test procedures may be of interest and use because it is, so to speak, an automatic procedure.

Matrix notation, methods, and results are used and assumed known to the reader. The notation, and such results of [20] as are needed, will be used without further summary herein.

2. Components of information. Since $I(1:2)$ and $J(1,2)$ are additive for independent random variables, for a random sample of n observations $I_n(1:2) = nI(1:2)$ and $J_n(1,2) = nJ(1,2)$, where $I(1:2)$ and $J(1,2)$ are given, respectively, by (2.8) and (2.7) of [20].

As is well known, the averages and the variances and covariances in samples from a multivariate normal population are independently distributed, respectively, in a normal distribution and in the Wishart distribution (see, for example, Wilks [33]). Computing the appropriate values from the respective distributions, it is readily found (see, for example, Hoyt [14]), that for the averages

$$
\begin{aligned}
I'(1:2;\bar{x}) &= \tfrac{1}{2}\log\frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} - \frac{k}{2} + \tfrac{1}{2}\operatorname{tr}\sigma_{(1)}\sigma_{(2)}^{-1} + \frac{n}{2}\delta'\sigma_{(2)}^{-1}\delta \\
&= \frac{n}{2}\delta'\sigma^{-1}\delta \qquad \text{for } \sigma_{(1)} = \sigma_{(2)} = \sigma,
\end{aligned} \tag{2.1}
$$

$$
\begin{aligned}
J'(1,2;\bar{x}) &= \tfrac{1}{2}\operatorname{tr}\left[(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})\right] + \frac{n}{2}\delta'(\sigma_{(1)}^{-1} + \sigma_{(2)}^{-1})\delta \\
&= n\delta'\sigma^{-1}\delta \qquad \text{for } \sigma_{(1)} = \sigma_{(2)} = \sigma,
\end{aligned} \tag{2.2}
$$

where $\delta = \mu_{(1)} - \mu_{(2)}$, and for the sample unbiased variances and covariances,

$$
I'(1:2;S) = \frac{n-1}{2}\left(\log\frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} - k + \operatorname{tr}\sigma_{(1)}\sigma_{(2)}^{-1}\right), \tag{2.3}
$$

$$
J'(1,2;S) = \frac{n-1}{2}\operatorname{tr}\left[(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})\right]. \tag{2.4}
$$

We thus have from the preceding,

$$nI(1:2) = I'(1:2;\bar{x}) + I'(1:2;S), \tag{2.5}$$

$$nJ(1,2) = J'(1,2;\bar{x}) + J'(1,2;S). \tag{2.6}$$

3. Estimates of information. The procedure we shall use (replacing population parameters in $I(1:2)$, $J(1,2)$ by unbiased estimates appropriate to the hypotheses) is based on a principle of maximizing information, as may be seen by the following heuristic discussion.

Suppose that $g_2(y)$ and $g^*(y)$ are densities, satisfying the conditions of Section 4 of [21], such that for given $g_2(y)$ we require

$$I^* = \int g^*(y) \log \frac{g^*(y)}{g_2(y)}\, d\gamma(y) \tag{3.1}$$

to be a maximum, subject to

$$\int g^*(y)\, d\gamma(y) = 1, \qquad \int yg^*(y)\, d\gamma(y) = a. \tag{3.2}$$

This is equivalent to maximizing

$$U = \int\left(g^*(y) \log \frac{g^*(y)}{g_2(y)} + kg^*(y) + lyg^*(y)\right) d\gamma(y), \tag{3.3}$$

where k and l are arbitrary constants to be determined. The usual variational procedures lead to

$$\delta U = 0 = \int \delta g^*(y)\left[\log \frac{g^*(y)}{g_2(y)} + 1 + k + ly\right] d\gamma(y) \tag{3.4}$$

or

$$\log \frac{g^*(y)}{g_2(y)} + 1 + k + ly = 0. \tag{3.5}$$

This means that

$$g^*(y) = e^{-1-k-ly} g_2(y) \tag{3.6}$$

or, by integration, that

$$1 = e^{-1-k} \int e^{-ly} g_2(y)\, d\gamma(y) = e^{-1-k} M_2(t), \tag{3.7}$$

where we have replaced $-l$ by t.

Thus,

$$g^*(y) = \frac{e^{ty} g_2(y)}{M_2(t)}, \qquad M_2(t) = \int e^{ty} g_2(y)\, d\gamma(y); \tag{3.8}$$

since this means that

$$(3.9)\qquad I^* = \int g^*(y) \log \frac{g^*(y)}{g_2(y)}\, d\gamma(y) = at - \log M_2(t),$$

the desired maximum occurs for that value of t which maximizes $at - \log M_2(t)$, or when $I^* = -\log m_2(a)$, in the notation of Section 4 of [21]. (It might be noted that $k \log m_2(a) = -kI^*$, where k is Boltzmann's constant, is the entropy of the distribution whose density is $g_2(y)$ (cf. Khinchin [18]).) Also,

$$(3.10)\qquad J^* = \int (g^*(y) - g_2(y)) \log \frac{g^*(y)}{g_2(y)}\, d\gamma(y) = t(a)(a - E_2(y)).$$

For a simple hypothesis, the parameters of $g_2(y)$ are completely specified. For a composite hypothesis, say $\theta \varepsilon \Theta$, where θ is a vector of the parameters and Θ is some subset of the parameter space, we use as the appropriate test of the null hypothesis (that relative to $g_2(y)$) the value given by

$$(3.11)\qquad \hat{I} = \min_{\theta \varepsilon \Theta} I^* = I^*(\hat{\theta})$$

and, correspondingly,

$$(3.12)\qquad \hat{J} = J^*(\hat{\theta}).$$

We shall carry through the foregoing in detail for some cases; in others, we shall apply the procedure of replacement of the parameters by unbiased estimates.

4. Single sample. Consider problem (a) of Section 1. For a sample of n observations from a multivariate normal population with mean matrix $\mu' = (\mu_1, \mu_2, \cdots, \mu_k)$ and covariance matrix σ, the moment-generating function of the sample averages $\bar{x}' = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$ and V_{ii} and $2V_{ij}$, $i \neq j$, the elements of the matrix $V = NS$, where N is the number of degrees of freedom and S is the sample unbiased covariance matrix, is known to be given by ([33], p. 121)

$$(4.1)\qquad M_2(t, T) = |I - 2\sigma T|^{-N/2} \exp\left(t'\mu + \tfrac{1}{2}t' \frac{\sigma}{n} t\right),$$

where $t' = (t_1, t_2, \cdots, t_k)$, $T = (t_{ij})$, $i, j = 1, 2, \cdots, k$.

We want to determine g^* of (3.9) (the conjugate distribution of Khinchin [18]) so as to have the observed unbiased estimates as its parameters, that is to say, $a = (\bar{x}, V)$, which means that we seek the values of t and T which will maximize (cf. [18], Section 33)

$$(4.2)\qquad I^* = t'\bar{x} + \operatorname{tr} TV - t'\mu - \tfrac{1}{2}t' \frac{\sigma}{n} t + \frac{N}{2} \log |I - 2\sigma T|.$$

Differentiating with respect to t and T (see [8], p. 364), we have

$$(4.3)\qquad \bar{x} - \mu - \frac{\sigma t}{n} = 0, \qquad \operatorname{tr} (dT)V - N \operatorname{tr} (I - 2\sigma T)^{-1}\sigma(dT) = 0,$$

from which are derived

$$t = n\sigma^{-1}(\bar{x} - \mu), \quad T = \tfrac{1}{2}\sigma^{-1} - \tfrac{1}{2}S^{-1}. \tag{4.4}$$

Using the values given by (4.4), (4.2) becomes

$$I^* = \frac{n}{2}(\bar{x} - \mu)'\sigma^{-1}(\bar{x} - \mu) + \frac{N}{2}(\log|\sigma| / |S| - k + \operatorname{tr} S\sigma^{-1}). \tag{4.5}$$

The null hypothesis specifies σ but not μ. It is clear that for variations of μ, I^* in (4.5) is a minimum for $\hat{\mu} = \bar{x}$, or

$$\hat{I} = \min_{\mu} I^* = I^*(\hat{\mu}) = \frac{N}{2}(\log|\sigma| / |S| - k + \operatorname{tr} S\sigma^{-1}). \tag{4.6}$$

Note that (4.6) is (2.3) with $\sigma_{(1)} = S$, $\sigma_{(2)} = \sigma$.

The case of a single sample of n observations from a k-variate normal population was considered in some detail by Hoyt [14]. Hoyt showed that asymptotically $2\hat{I}$ has a chi-square distribution with $k(k + 1)/2$ d.f., and to a closer approximation, R. A. Fisher's B distribution.

In considering tests of significance in factor analysis, Bartlett [4], using a "homogeneous" likelihood function, and Rippe [26], using the likelihood-ratio test procedure for the problem of tests of significance of components in matrix factorization, arrived at the statistic $2\hat{I}$ and the same conclusion as to its asymptotic chi-square distribution.

For the hypothesis of independence of variates, i.e.,

$$\sigma = (\sigma_{ij}), \qquad \sigma_{ij} = 0, \qquad i \neq j, \tag{4.7}$$

we may write (4.6) as

$$2\hat{I} = -N \log|R| + N\left[\sum_{i=1}^{k}\left(\frac{S_{ii}}{\sigma_{ii}} + \log\frac{\sigma_{ii}}{S_{ii}} - 1\right)\right], \tag{4.8}$$

where R is the matrix of sample correlation coefficients, or

$$2\hat{I} = 2\hat{I}'_R + 2\hat{I}'_S, \tag{4.9}$$

where $2\hat{I}'_R = -N \log|R|$ (cf. [20], p. 94). Wilks [31] has shown that when (4.7) holds, the s_{ii} and r_{ij} are independent, so that $\hat{I}'_R$ and $\hat{I}'_S$ are independent. It follows from the discussion of Section 9 that, asymptotically, $2\hat{I}'_R$ has a chi-square distribution with $k(k - 1)/2$ d.f., and $2\hat{I}'_S$ has a chi-square distribution with k d.f. It is shown in Section 9 that a better approximation to the distribution of $2\hat{I}'_R$ is given by Fisher's B distribution ([10], p. 14.665) with

$$\beta^2 = k(k - 1)(2k + 5)/12N, \quad B^2 = 2\hat{I}'_R, \quad n_1 = k(k - 1)/2.$$

5. Homogeneity of means. Consider problem (b) of Section 1. We will first discuss the case for two samples. Suppose we have two independent samples,

having, respectively, n_1, n_2 independent observations from k-variate normal populations with respective covariance matrices Σ_1, Σ_2. We want to test the null hypothesis that the population mean vectors are equal, i.e.,

$$H_2: \mu_{(1)} = \mu_{(2)} = \mu, \qquad \Sigma_1, \Sigma_2, \tag{5.1}$$

with no specification about Σ_1 and Σ_2.

Using the notation already introduced in Section 4, we want to determine g^*, with $a = (\bar{x}_{(1)}, \bar{x}_{(2)}, V_1, V_2)$, which means that we seek the values of $t_{(i)}$, T_i, $i = 1, 2$, which will maximize

$$\begin{aligned} I^* = t'_{(1)}\bar{x}_{(1)} - t'_{(1)}\mu - \tfrac{1}{2}t'_{(1)}\frac{\Sigma_1}{n_1}t_{(1)} + \operatorname{tr} T_1V_1 + \frac{N_1}{2}\log|I - 2\Sigma_1T_1| \\ + t'_{(2)}\bar{x}_{(2)} - t'_{(2)}\mu - \tfrac{1}{2}t'_{(2)}\frac{\Sigma_2}{n_2}t_{(2)} + \operatorname{tr} T_2V_2 + \frac{N_2}{2}\log|I - 2\Sigma_2T_2|. \end{aligned} \tag{5.2}$$

Following the procedure as used for (4.3), we find that the sought for values are given by

$$\begin{aligned} t_{(1)} &= n_1\Sigma_1^{-1}(\bar{x}_{(1)} - \mu), & t_2 &= n_2\Sigma_2^{-1}(\bar{x}_{(2)} - \mu), \\ T_1 &= \tfrac{1}{2}\Sigma_1^{-1} - \tfrac{1}{2}S_1^{-1}, & T_2 &= \tfrac{1}{2}\Sigma_2^{-1} - \tfrac{1}{2}S_2^{-1}, \end{aligned} \tag{5.3}$$

for which values I^* of (5.2) becomes

$$\begin{aligned} I^* = \frac{n_1}{2}(\bar{x}_{(1)} - \mu)'\Sigma_1^{-1}(\bar{x}_{(1)} - \mu) + \frac{n_2}{2}(\bar{x}_{(2)} - \mu)'\Sigma_2^{-1}(\bar{x}_{(2)} - \mu) \\ + \frac{N_1}{2}\left(\log\frac{|\Sigma_1|}{|S_1|} - k + \operatorname{tr} S_1\Sigma_1^{-1}\right) + \frac{N_2}{2}\left(\log\frac{|\Sigma_2|}{|S_2|} - k + \operatorname{tr} S_2\Sigma_2^{-1}\right). \end{aligned} \tag{5.4}$$

The null hypothesis requires equality of the means with no specification on the covariance matrices. It is clear that for variations of Σ_1 and Σ_2, I^* will be a minimum for $\hat{\Sigma}_1 = S_1$, $\hat{\Sigma}_2 = S_2$, and for $\hat{\mu}$ satisfying

$$0 = n_1S_1^{-1}(\bar{x}_{(1)} - \hat{\mu}) + n_2S_2^{-1}(\bar{x}_{(2)} - \hat{\mu}) \tag{5.5}$$

or

$$\hat{\mu} = (n_1S_1^{-1} + n_2S_2^{-1})^{-1}(n_1S_1^{-1}\bar{x}_{(1)} + n_2S_2^{-1}\bar{x}_{(2)}). \tag{5.6}$$

For convenience let $d = \bar{x}_{(1)} - \bar{x}_{(2)}$, $A = n_1S_1^{-1}$, $B = n_2S_2^{-1}$; substituting in (5.4) we get

$$\begin{aligned} &2\hat{I}(\hat{\mu}, \hat{\Sigma}_1, \hat{\Sigma}_2) \\ &\quad = \operatorname{tr}\{[B(A + B)^{-1}A(A + B)^{-1}B + A(A + B)^{-1}B(A + B)^{-1}A]\,dd'\}. \end{aligned} \tag{5.7}$$

But $B(A + B)^{-1}A = [A^{-1}(A + B)B^{-1}]^{-1} = (B^{-1} + A^{-1})^{-1}$ and $A(A + B)^{-1}B = [B^{-1}(A + B)A^{-1}]^{-1} = (B^{-1} + A^{-1})^{-1}$, so that finally

$$
\begin{aligned}
2\hat{I} &= \operatorname{tr}[(B^{-1} + A^{-1})^{-1} dd'] \\
&= d'(B^{-1} + A^{-1})^{-1} d \qquad (5.8) \\
&= (\bar{x}_{(1)} - \bar{x}_{(2)})' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}).
\end{aligned}
$$

It is readily found that in this case $\hat{J} = 2\hat{I}$.

For the single variate case, see Fisher [10], pp. 35.174–35.180.

Linear discriminant function. Consider $y = \alpha' x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$, the same linear compound for each sample. Since y is normally distributed, we seek α so as to maximize

$$
(5.9) \qquad 2\hat{I}'(\hat{\mu}, \hat{\Sigma}_1, \hat{\Sigma}_2 ; y) = \frac{\alpha' dd' \alpha}{\alpha' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right) \alpha}.
$$

As is easily determined (cf. [20], p. 91), the maximum occurs for

$$
\alpha = \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} d \qquad \text{and} \qquad 2\hat{I}'(y) = 2\hat{I}.
$$

r-samples. Suppose we have r independent samples, having, respectively, n_i, $i = 1, 2, \cdots, r$, independent observations each, from k-variate normal populations with respective covariance matrices Σ_i, $i = 1, 2, \cdots, r$, and we want to test the null hypothesis that the population mean vectors are equal, i.e.,

$$
(5.10) \qquad H_2 : \mu_{(1)} = \mu_{(2)} = \cdots = \mu_{(r)} = \mu, \qquad \Sigma_1, \Sigma_2, \cdots, \Sigma_r,
$$

with no specification about the Σ_i.

Without repeating the details, we find, in this case, that

$$
(5.11) \quad I^* = \sum_{i=1}^{r} \frac{n_i}{2} (\bar{x}_{(i)} - \mu)' \Sigma_i^{-1} (\bar{x}_{(i)} - \mu) + \sum_{i=1}^{r} \frac{N_i}{2} \left(\log \frac{|\Sigma_i|}{|S_i|} - k + \operatorname{tr} S_i \Sigma_i^{-1}\right),
$$

$$
(5.12) \qquad \hat{\Sigma}_i = S_i, \qquad \hat{\mu} = \left(\sum_{i=1}^{r} n_i S_i^{-1}\right)^{-1} \left(\sum_{i=1}^{r} n_i S_i^{-1} \bar{x}_{(i)}\right) = \bar{x}.
$$

$$
(5.13) \qquad 2\hat{I} = \sum_{i=1}^{r} n_i (\bar{x}_{(i)} - \bar{x})' S_i^{-1} (\bar{x}_{(i)} - \bar{x}).
$$

On the null hypothesis, $2\hat{I}$ has an asymptotic chi-square distribution with $(r - 1)k$ d.f.

Covariance matrices equal. If we assume that the population covariance matrices are equal, i.e., that $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_r = \Sigma$, and want to test the null hypothesis that the population mean vectors are equal, then, without repeating the details, we find, in this case, that

$$
(5.14) \quad I^* = \sum_{i=1}^{r} \frac{n_i}{2} (\bar{x}_{(i)} - \mu)' \Sigma^{-1} (\bar{x}_{(i)} - \mu) + \frac{N}{2} \left(\log \frac{|\Sigma|}{|S|} - k + \operatorname{tr} S \Sigma^{-1}\right),
$$

where $NS = N_1S_1 + \cdots + N_rS_r$, $N = N_1 + N_2 + \cdots + N_r$, and that

$$\hat{\Sigma} = S, \qquad n\hat{\mu} = n\bar{x} = n_1\bar{x}_1 + \cdots + n_r\bar{x}_r, \quad n = n_1 + n_2 + \cdots + n_r, \tag{5.15}$$

$$\begin{aligned} 2\hat{I} &= \sum_{i=1}^{r} n_i(\bar{x}_{(i)} - \bar{x})'S^{-1}(\bar{x}_{(i)} - \bar{x}) \\ &= \mathrm{tr}\, S^{-1}(n_1\, d_{(1)}\, d'_{(1)} + \cdots + n_r\, d_{(r)}\, d'_{(r)}) \\ &= \mathrm{tr}\, S^{-1}\, S^*, \end{aligned} \tag{5.16}$$

where $d_{(i)} = \bar{x}_{(i)} - \bar{x}$, $S^* = \sum_{i=1}^{r} n_i\, d_{(i)}\, d'_{(i)}$ (cf. Hotelling [13]). It is readily found that in this case $\hat{J} = 2\hat{I}$.

Asymptotically, $2\hat{I} = \hat{J}$ has a chi-square distribution with $k(r-1)$ d.f. on the null hypothesis (cf. [25], p. 372). This will be shown in Section 10.

Linear discriminant function. Consider $y = \alpha'x = \alpha_1x_1 + \alpha_2x_2 + \cdots + \alpha_kx_k$, the same linear compound for each sample. Since y is normally distributed, we have for the y's, using (5.16),

$$\begin{aligned} 2\hat{I}'(y) &= \frac{n_1(\alpha'\, d_{(1)})^2 + \cdots + n_r(\alpha'\, d_{(r)})^2}{\alpha'S\alpha} \\ &= \frac{\alpha'(n_1\, d_{(1)}\, d'_{(1)} + \cdots + n_r\, d_{(r)}\, d'_{(r)})\alpha}{\alpha'S\alpha} \\ &= \frac{\alpha'S^*\alpha}{\alpha'S\alpha}, \end{aligned} \tag{5.17}$$

with the symbols as defined in (5.16). For the linear compound for which $2\hat{I}'(y)$ is a maximum, the usual calculus procedures yield the result that the α's must satisfy $S^*\alpha = lS\alpha$, where l is the largest root of the equation $|S^* - lS| = 0$, which has (almost everywhere) p positive and $(k-p)$ zero roots, where $p \leq \min(k, r-1)$ (cf. [28]). If we denote the positive roots in descending order as $l_1, l_2, \cdots, l_p$,

$$\begin{aligned} 2\hat{I} = \hat{J} &= \mathrm{tr}\, S^{-1}S^* = l_1 + l_2 + \cdots + l_p \\ &= \hat{J}'(l_1) + \hat{J}'(l_2) + \cdots + \hat{J}'(l_p). \end{aligned} \tag{5.18}$$

The discrimination efficiency of the linear compound associated with l_i is given by

$$\mathrm{Eff} = \frac{\hat{J}'(l_i)}{\hat{J}} = \frac{l_i}{l_1 + l_2 + \cdots + l_p}. \tag{5.19}$$

In this case, asymptotically we have, on the null hypothesis, the chi-square decomposition (cf. [25], p. 373)

$$\begin{array}{ll} \hat{J}'(l_p) = l_p, & |k - (r-1)| + 1 \ \text{d.f.} \\ \hat{J}'(l_{p-1}) = l_{p-1}, & |k - (r-1)| + 3 \ \text{d.f.} \\ \cdots \quad \cdots & \cdots \\ \hline \hat{J} = l_1 + l_2 + \cdots + l_p, & k(r-1) \ \text{d.f.} \end{array}$$

This is to be taken in the sense that $l_{m+1} + \cdots + l_p$ is asymptotically a chi-square, not that $l_{m+1}, \cdots, l_p$ have asymptotic independent chi-square distributions (see (10.3)).

EXAMPLE. Consider the following data from a problem discussed by Bartlett and which was also considered in [20], p. 93. (See further references therein.) Here, $r = 8$, $k = 2$, $n = n_1 + \cdots + n_8 = 57$,

$$49S = \begin{pmatrix} 136{,}972.6 & 58{,}549.0 \\ 58{,}549.0 & 71{,}496.1 \end{pmatrix}, \qquad S^* = \begin{pmatrix} 12{,}496.8 & -6{,}786.6 \\ -6{,}786.6 & 32{,}985.0 \end{pmatrix},$$

and the roots of $|S^* - lS| = 0$ are given by $l_1 = 44.68667$, $l_2 = 3.09106$. Also,

$$\begin{aligned} \hat{J}'(l_2) &= 3.09106 \quad 6 \text{ d.f.} \\ \hat{J}'(l_1) &= 44.68667 \quad 8 \text{ d.f.} \\ \hline \hat{J} &= 47.77773 \quad 14 \text{ d.f.} \end{aligned}$$

Since only $\hat{J}'(l_1)$ is significant, the linear discriminant function $y = x_2 - 0.535x_1$, associated with l_1, is affected by the treatments and is practically sufficient.

6. Multivariate linear hypothesis. Consider problem (c) of Section 1. Let $Z_{(i)} = Y_{(i)} - BX_{(i)}$, $i = 1, 2, \cdots, n$, where $Z'_{(i)} = (Z_{i1}, \cdots, Z_{ik_2})$, $Y'_{(i)} = (y_{i1}, \cdots, y_{ik_2})$, $X'_{(i)} = (x_{i1}, \cdots, x_{ik_1})$, $B = (\beta_{rs})$, $r = 1, 2, \cdots, k_2$, $s = 1, 2, \cdots, k_1$, $k_1 \geq k_2$, and the $Z_{(i)}$ are independent k_2-variate normal random vectors with zero means and common covariance matrix Σ. The $Y_{(i)}$ are stochastic and the $X_{(i)}$ are considered known. The usual unbiased estimate of B is given by (see [2], pp. 103–104) $\hat{B} = (Y'X)(X'X)^{-1}$, where

$$Y' = (Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}), \quad X' = X_{(1)}, X_{(2)}, \cdots, X_{(n)}),$$

and that of Σ is given by $(n - k_1)\hat{\Sigma} = \hat{Z}'\hat{Z} = (Y' - \hat{B}X')(Y - X\hat{B}') = Y'Y - \hat{B}X'X\hat{B}' = Y'Y - (Y'X)(X'X)^{-1}(X'Y)$.

Let us now consider the hypotheses

$$\text{(6.1)} \qquad \begin{aligned} &H_1: E_1(Y_{(i)}) = BX_{(i)}, && i = 1, 2, \cdots, n, \\ &H_2: E_2(Y_{(i)}) = 0, && \text{i.e., } B = 0. \end{aligned}$$

As in pp. 90–91 of [20], we have

$$2I(1:2) = J(1, 2) = \delta'\sigma^{-1}\delta$$

$$\text{(6.2)} \qquad = (X'_{(1)}B', X'_{(2)}B', \cdots, X'_{(n)}B') \begin{bmatrix} \Sigma^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Sigma^{-1} \end{bmatrix} \begin{bmatrix} BX_{(1)} \\ BX_{(2)} \\ \vdots \\ BX_{(n)} \end{bmatrix}$$

$$= X'_{(1)}B'\Sigma^{-1}BX_{(1)} + \cdots + X'_{(n)}B'\Sigma^{-1}BX_{(n)}$$
$$= \operatorname{tr} \Sigma^{-1}B(X_{(1)}X'_{(1)} + \cdots + X_{(n)}X'_{(n)})B'$$
$$= \operatorname{tr} \Sigma^{-1}BX'XB'.$$

Using the estimates given above, we get as the estimate of $J(1, 2)$ (cf. [20], p. 96)

$$\begin{aligned} 2\hat{I}(1:2) &= \hat{J}(1, 2) = \operatorname{tr} \hat{\Sigma}^{-1}\hat{B}X'X\hat{B}' \\ &= (n - k_1) \operatorname{tr} (Y'Y - (Y'X)(X'X)^{-1}(X'Y))^{-1}(Y'X)(X'X)^{-1}(X'Y) \qquad (6.3)\\ &= (n - k_1) \operatorname{tr} S_{22.1}^{-1}S_{21}S_{11}^{-1}S_{12}, \end{aligned}$$

where $X'X = nS_{11}$, $X'Y = nS_{12}$, $Y'X = nS_{21}$, $Y'Y = nS_{22}$, and

$$S_{22.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}.$$

We may also express $\hat{J}(1, 2)$ as $(n - k_1)$ times the sum of the k_2 roots (almost everywhere positive) of the determinantal equation $|S_{21}S_{11}^{-1}S_{12} - lS_{22.1}| = 0$.

As in the preceding section,

$$2\hat{I}(1:2) = \hat{J}(1, 2) = (n - k_1)(l_1 + l_2 + \cdots + l_{k_2}), \qquad (6.4)$$

asymptotically on the null hypothesis H_2, has a chi-square distribution with k_1k_2 d.f. (see Section 10). By replacing $S_{22.1}$ by its value as given above,

$$|S_{21}S_{11}^{-1}S_{12} - lS_{22.1}| = 0 = |S_{21}S_{11}^{-1}S_{12} - r^2S_{22}|, \text{ where } l = r^2/(1 - r^2).$$

The r's thus defined are Hotelling's canonical correlation coefficients. (See [20], pp. 95–99, and further references therein.) We may also write (6.4) as (cf. [20], p. 97)

$$2\hat{I}(1:2) = \hat{J}(1,2) = (n - k_1)\left(\frac{r_1^2}{1 - r_1^2} + \frac{r_2^2}{1 - r_2^2} + \cdots + \frac{r_{k_2}^2}{1 - r_{k_2}^2}\right). \qquad (6.5)$$

On the null hypothesis $B = 0$, the results are equivalent to those for the null hypothesis that in a k-variate normal population, the set of the first k_1 variates is uncorrelated with the set of the last k_2 variates, $k = k_1 + k_2$. The latter hypothesis is the one considered in [20], pp. 95–99.

Linear discriminant function. For the problem of this section, consider $w_i = \alpha'Y_{(i)} = \alpha_1Y_{i1} + \alpha_2Y_{i2} + \cdots + \alpha_{k_2}Y_{ik_2}$, $i = 1, 2, \cdots, n$, the same linear compound of the y's for each observation. Since the w's are normally distributed with $\sigma_{w_i}^2 = \alpha'\Sigma\alpha$, we have for the w's

$$\begin{aligned} 2I'(1:2; w) = J'(1,2; w) &= \frac{(\alpha'BX_{(1)})^2 + \cdots + (\alpha'BX_{(n)})^2}{\alpha'\Sigma\alpha} \\ &= \frac{\alpha'B(X_{(1)}X'_{(1)} + \cdots + X_{(n)}X'_{(n)})B'\alpha}{\alpha'\Sigma\alpha} \qquad (6.6)\\ &= \frac{\alpha'BX'XB'\alpha}{\alpha'\Sigma\alpha}. \end{aligned}$$

To find the linear compound for which $J'(1,2;w)$ is a maximum, the usual calculus procedures yield the result that the α's must satisfy $BX'XB'\alpha = \lambda\Sigma\alpha$, where λ is the largest root of the equation $|BX'XB' - \lambda\Sigma| = 0$. Denoting the k_2 positive roots in descending order as $\lambda_1, \lambda_2, \cdots, \lambda_{k_2}$,

$$(6.7)\quad \begin{aligned} 2I(1:2) = J(1,2) &= \operatorname{tr} \Sigma^{-1}BX'XB' = \lambda_1 + \lambda_2 + \cdots + \lambda_{k_2} \\ &= J'(1,2;\lambda_1) + \cdots + J'(1,2:\lambda_{k_2}). \end{aligned}$$

Using the estimates as in (6.3) and (6.4), we have

$$(6.8)\quad \begin{aligned} 2\hat{I}(1:2) = \hat{J}(1,2) &= (n-k_1)(l_1 + l_2 + \cdots + l_{k_2}) \\ &= \hat{J}'(1,2;l_1) + \hat{J}'(1,2;l_2) + \cdots + \hat{J}'(1,2;l_{k_2}). \end{aligned}$$

The canonical correlations enter as before.

In this case, too, asymptotically on the null hypothesis H_2, we have the chi-square decomposition (see Section 10)

$$\begin{array}{ll}
\hat{J}'(1,2;l_{k_2}) = (n-k_1)l_{k_2} = (n-k_1)r_{k_2}^2/(1-r_{k_2}^2) & k_1 - k_2 + 1 \text{ d.f.} \\
\hat{J}'(1,2;l_{k_2-1}) = (n-k_1)l_{k_2-1} = (n-k_1)r_{k_2-1}^2/(1-r_{k_2-1}^2) & k_1 - k_2 + 3 \text{ d.f.} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots \\
\hat{J}'(1,2;l_1) = (n-k_1)l_1 = (n-k_1)r_1^2/(1-r_1^2) & k_1 + k_2 - 1 \text{ d.f.} \\
\hline
\hat{J}(1,2) = (n-k_1)\sum_{i=1}^{k_2} l_i = (n-k_1)\sum_{i=1}^{k_2} r_i^2 \Big/ (1-r_i^2) & k_1k_2 \text{ d.f.}
\end{array}$$

This is to be taken in the sense that $(n-k_1)(l_{m+1} + \cdots + l_{k_2})$ is asymptotically a chi-square, not that $(n-k_1)l_{m+1}, \cdots, (n-k_1)l_{k_2}$ have asymptotic independent chi-square distributions. (See (10.4).)

Example. By way of illustration, we use the data already discussed by Hotelling and the values derived in [20], p. 98, where it was found that $r_1^2 = .1556$, $r_2^2 = .0047$, $r_1^2/(1-r_1^2) = .1843$, $r_2^2/(1-r_2^2) = .0047$; and since $n - k_1 = 139 - 2 = 137$ (there were 140 observations but the values were computed about the sample averages),

$$\begin{array}{rl}
\hat{J}'(1,2;r_2) = & .6439 \ 1 \text{ d.f.} \\
\hat{J}'(1,2;r_1) = & 25.2491 \ 3 \text{ d.f.} \\
\hline
\hat{J}(1,2) = & 25.8930 \ 4 \text{ d.f.}
\end{array}$$

Since only $\hat{J}'(1,2;r_1)$ is significant, the linear discriminant function associated with r_1, $w = -2.4404\, y_1 + y_2$, is the only such linear function and is practically sufficient, confirming the inference made in [20], p. 99.

Subhypothesis. We return to the problem at the beginning of this section and separate the k_1 x's into two sets of q_1 and q_2, $k_1 = q_1 + q_2$. With a corresponding partition of the matrix B, we now have $Z_{(i)} = Y_{(i)} - CX_{(1i)} - DX_{(2i)}$, where $X_{(i)} = \begin{pmatrix} X_{(1i)} \\ X_{(2i)} \end{pmatrix}$; $B = (C, D)$, where C and D are, respectively, $k_2 \times q_1$, $k_2 \times q_2$

matrices; or $Z' = Y' - CX_1' - DX_2'$, with Z and Y as previously defined and

$$X' = (X_{(1)}, X_{(2)}, \cdots, X_{(n)}) = \left(\begin{pmatrix} X_{(11)} \\ X_{(21)} \end{pmatrix}, \cdots, \begin{pmatrix} X_{(1n)} \\ X_{(2n)} \end{pmatrix}\right) = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix}.$$

With the same assumptions as to the $Z_{(i)}$, we now consider the hypotheses

$$\begin{aligned} H_1: E_1(Y_{(i)}) &= C_1X_{(1i)} + D_1X_{(2i)}, \\ H_2: E_2(Y_{(i)}) &= C_2X_{(1i)} + D_2X_{(2i)}, \end{aligned} \qquad i = 1, 2, \cdots, n. \tag{6.9}$$

Applying the same procedures as previously, it is found that now

$$\begin{aligned} 2I(1:2) &= J(1,2) \\ &= \operatorname{tr} \Sigma^{-1}\left\{((C_1 - C_2), (D_1 - D_2))\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{pmatrix} (C_1 - C_2)' \\ (D_1 - D_2)' \end{pmatrix}\right\}, \end{aligned} \tag{6.10}$$

where

$$X'X = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix}(X_1X_2) = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}.$$

In particular, we wish to test the null hypothesis that $D_2 = 0$. For C_1 and D_1, the estimation procedure previously used for B, [2], yields here

$$(\hat{C}_1, \hat{D}_1)\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = (Y'X_1, Y'X_2), \tag{6.11}$$

or

$$\begin{aligned} \hat{C}_1S_{11} + \hat{D}_1S_{21} &= Y'X_1, \\ \hat{C}_1S_{12} + \hat{D}_1S_{22} &= Y'X_2; \end{aligned} \tag{6.12}$$

and for C_2,

$$\hat{C}_2S_{11} = Y'X_1. \tag{6.13}$$

From (6.12) it is readily found that

$$\hat{D}_1 = Y'X_{2.1}S_{22.1}^{-1}, \qquad \hat{C}_1 = Y'X_1S_{11}^{-1} - \hat{D}_1S_{21}S_{11}^{-1}, \tag{6.14}$$

where $X_{2.1} = X_2 - X_1S_{11}^{-1}S_{12}$, $S_{22.1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$.

For the estimate of Σ, we have as before

$$(n - k_1)\hat{\Sigma} = Y'Y - (\hat{C}_1, \hat{D}_1)\begin{pmatrix} S_{11}S_{12} \\ S_{21}S_{22} \end{pmatrix}\begin{pmatrix} \hat{C}_1' \\ \hat{D}_1' \end{pmatrix}; \tag{6.15}$$

and for the estimate of $J(1,2)$,

$$\hat{J}(1,2) = \operatorname{tr} \hat{\Sigma}^{-1}\left\{((\hat{C}_1 - \hat{C}_2), \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix} (\hat{C}_1 - \hat{C}_2)' \\ \hat{D}_1' \end{pmatrix}\right\}. \tag{6.16}$$

Using the values given in (6.13) and (6.14), it is found that

$$\begin{aligned} (\hat{C}_1, \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix} \hat{C}_1' \\ \hat{D}_1' \end{pmatrix} &= \hat{C}_2 S_{11} \hat{C}_2' + \hat{D}_1 S_{22.1} \hat{D}_1' \\ &= Y'X_1 S_{11}^{-1} X_1' Y + Y' X_{2.1} S_{22.1}^{-1} X_{2.1}' Y, \\ ((\hat{C}_1 - \hat{C}_2), \hat{D}_1)\begin{pmatrix} S_{11} S_{12} \\ S_{21} S_{22}\end{pmatrix}\begin{pmatrix} (\hat{C}_1 - \hat{C}_2)' \\ \hat{D}_1' \end{pmatrix} &= \hat{D}_1 S_{22.1} \hat{D}_1'. \end{aligned} \tag{6.17}$$

It is readily verified that

$$X_1 S_{11}^{-1} X_1' X_{2.1} S_{22.1}^{-1} X_{2.1}' = 0\,, \tag{6.18}$$

and since $X_{2.1}' X_{2.1} = S_{22.1}$,

$$(I - X_1 S_{11}^{-1} X_1' - X_{2.1} S_{22.1}^{-1} X_{2.1}') X_{2.1} S_{22.1}^{-1} X_{2.1}' = 0; \tag{6.19}$$

that is to say, the two factors in $\hat{J}(1,2)$ are independent.

These results are summarized in Tables 1 and 2.

We omit a discussion of linear discriminant functions for this case.

7. Homogeneity of covariance matrices. Consider problem (d) of Section 1. For its special interest, we consider first the case of two samples and then the general case.

(7.1) *Two samples.* Suppose that we have two independent samples with n_1

TABLE 1

Due to	d.f.	Generalized sum of squares
$\hat{C}_2$	q_1	$\hat{C}_2 S_{11} \hat{C}_2' = Y'X_1 S_{11}^{-1} X_1' Y$
Difference	q_2	$\hat{D}_1 S_{22.1} \hat{D}_1' = Y' X_{2.1} S_{22.1}^{-1} X_{2.1}' Y$
$\hat{C}_1, \hat{D}_1$	k_1	$\hat{B} X' X \hat{B}' = Y' X_1 S_{11}^{-1} X_1' Y + Y' X_{2.1} S_{22.1}^{-1} X_{2.1}' Y$
Difference	$n - k_1$	$Y'Y - \hat{B}X'X\hat{B}' = (n - k_1)\hat{\Sigma}$
Total	n	$Y'Y$

TABLE 2

Test	Asymptotic distribution on the	Null hypothesis
$\operatorname{tr} \hat{\Sigma}^{-1} \hat{B} X' X \hat{B}'$	chi-square $k_1 k_2$ d.f.	$B = 0$; i.e., $C_2 = 0$, $D_2 = 0$
$\operatorname{tr} \hat{\Sigma}^{-1} \hat{D}_1 S_{22.1} \hat{D}_1'$	chi-square $q_2 k_2$ d.f.	$D_2 = 0$

and n_2 independent observations, respectively, from k-variate normal populations for which we make no specification about the means, and suppose that for the population covariance matrices we have the two hypotheses, $H_1\colon \Sigma_1 \neq \Sigma_2$ and $H_2\colon \Sigma_1 = \Sigma_2 = \Sigma$.

Using the notation already introduced in Section 4, we want to determine g^* with $a = (\bar{x}_{(1)}, \bar{x}_{(2)}, V_1, V_2)$, which means that we seek the values of $t_{(i)}$, T_i, $i = 1,2$, which will maximize (cf. (5.2))

$$\begin{aligned} I^* &= t'_{(1)}\bar{x}_{(1)} - t'_{(1)}\mu_{(1)} - \tfrac{1}{2}t'_{(1)}\frac{\Sigma}{n_1}t_{(1)} + \operatorname{tr} T_1V_1 + \frac{N_1}{2}\log|I - 2\Sigma T_1| \\ &\quad + t'_{(2)}\bar{x}_{(2)} - t'_{(2)}\mu_{(2)} - \tfrac{1}{2}t'_{(2)}\frac{\Sigma}{n_2}t_{(2)} + \operatorname{tr} T_2V_2 + \frac{N_2}{2}\log|I - 2\Sigma T_2|. \end{aligned} \tag{7.1.1}$$

Following the procedure as used for (4.3), we find that the sought-for values are given by (cf. (5.3))

$$\begin{aligned} t_{(1)} &= n_1\Sigma^{-1}(\bar{x}_{(1)} - \mu_{(1)}), & t_{(2)} &= n_2\Sigma^{-1}(\bar{x}_{(2)} - \mu_{(2)}), \\ T_1 &= \tfrac{1}{2}\Sigma^{-1} - \tfrac{1}{2}S_1^{-1}, & T_2 &= \tfrac{1}{2}\Sigma^{-1} - \tfrac{1}{2}S_2^{-1}, \end{aligned} \tag{7.1.2}$$

for which values I^* of (7.1.1) becomes

$$\begin{aligned} I^* &= \frac{n_1}{2}(\bar{x}_{(1)} - \mu_{(1)})'\Sigma^{-1}(\bar{x}_{(1)} - \mu_{(1)}) + \frac{n_2}{2}(\bar{x}_{(2)} - \mu_{(2)})'\Sigma^{-1}(\bar{x}_{(2)} - \mu_{(2)}) \\ &\quad + \frac{N_1}{2}\left(\log\frac{|\Sigma|}{|S_1|} - k + \operatorname{tr} S_1\Sigma^{-1}\right) + \frac{N_2}{2}\left(\log\frac{|\Sigma|}{|S_2|} - k + \operatorname{tr} S_2\Sigma^{-1}\right). \end{aligned} \tag{7.1.3}$$

For variations of $\mu_{(1)}$, $\mu_{(2)}$, and Σ, I^* will be a minimum for $\hat{\mu}_{(1)}$, $\hat{\mu}_{(2)}$, and $\hat{\Sigma}$ satisfying (see [8])

$$\begin{aligned} & n_1\hat{\Sigma}^{-1}(\bar{x}_{(1)} - \hat{\mu}_{(1)}) = 0, \qquad n_2\hat{\Sigma}^{-1}(\bar{x}_{(2)} - \hat{\mu}_{(2)}) = 0, \\ & 0 = -\frac{n_1}{2}(\bar{x}_{(1)} - \hat{\mu}_{(1)})'\hat{\Sigma}^{-1}\,d\Sigma\,\hat{\Sigma}^{-1}(\bar{x}_{(1)} - \hat{\mu}_{(1)}) \\ & \quad - \frac{n_2}{2}(\bar{x}_{(2)} - \hat{\mu}_{(2)})'\hat{\Sigma}^{-1}\,d\Sigma\,\hat{\Sigma}^{-1}(x_{(2)} - \hat{\mu}_{(2)}) + \frac{N_1}{2}\operatorname{tr}\hat{\Sigma}^{-1}\,d\Sigma \\ & \quad - \frac{N_1}{2}\operatorname{tr} S_1\hat{\Sigma}^{-1}\,d\Sigma\,\hat{\Sigma}^{-1} + \frac{N_2}{2}\operatorname{tr}\hat{\Sigma}^{-1}\,d\Sigma - \frac{N_2}{2}\operatorname{tr} S_2\hat{\Sigma}^{-1}\,d\Sigma\,\hat{\Sigma}^{-1}, \end{aligned} \tag{7.1.4}$$

from which we find that

$$\hat{\mu}_{(1)} = \bar{x}_{(1)}, \qquad \hat{\mu}_{(2)} = \bar{x}_{(2)}, \qquad (N_1 + N_2)\hat{\Sigma} = N_1S_1 + N_2S_2 = NS, \tag{7.1.5}$$

where $N = N_1 + N_2$; consequently (cf. Wilks [31], p. 489),

$$2\hat{I} = N_1\log\frac{|S|}{|S_1|} + N_2\log\frac{|S|}{|S_2|}. \tag{7.1.6}$$

It is readily found that the corresponding $\hat{J}$ is given by (cf. [20], p. 91)

$$\hat{J} = \frac{N_1 N_2}{2(N_1 + N_2)} (\operatorname{tr} S_1 S_2^{-1} + \operatorname{tr} S_2 S_1^{-1} - 2k). \tag{7.1.7}$$

It will be shown in Section 8 that $2\hat{I}$, for large N_1 and N_2, on the null hypothesis H_2, has a chi-square distribution with $k(k + 1)/2$ d.f., and to a better approximation, a non-central chi-square distribution, R. A. Fisher's B distribution.

Linear discriminant function. We seek a linear compound, the same for both samples, $y = \alpha' x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$, which will maximize (see (7.1.7))

$$\hat{J}'(y) = \frac{N_1 N_2}{2(N_1 + N_2)} \left(\frac{\alpha' S_1 \alpha}{\alpha' S_2 \alpha} + \frac{\alpha' S_2 \alpha}{\alpha' S_1 \alpha} - 2 \right). \tag{7.1.8}$$

The usual calculus procedures yield the result that α is obtained as a solution of $S_1 \alpha = F S_2 \alpha$, where F is a root of the determinantal equation $|S_1 - F S_2| = |N_1 S_1 - l N_2 S_2| = 0$, and $F = N_2 l / N_1$. It is found that the same linear function results from maximizing (see 7.1.6))

$$\hat{I}'(y) = \frac{N_1}{2} \log \frac{\alpha' S \alpha}{\alpha' S_1 \alpha} + \frac{N_2}{2} \log \frac{\alpha' S \alpha}{\alpha' S_2 \alpha}. \tag{7.1.9}$$

If the roots of the determinantal equation, which are almost everywhere positive, are $F_1, F_2, \cdots, F_k$ arranged in ascending order, then, as was shown in [20], Section 5, the maximum of $\hat{J}'(y)$ occurs for the linear compound associated with F_1 or F_k according as $F_1 F_k < 1$ or $F_1 F_k > 1$.

It may also be shown, readily, that

$$\begin{aligned} \hat{I} &= \hat{I}'(l_1) + \hat{I}'(l_2) + \cdots + \hat{I}'(l_k), \\ \hat{J} &= \hat{J}'(F_1) + \hat{J}'(F_2) + \cdots + \hat{J}'(F_k), \end{aligned} \tag{7.1.10}$$

where

$$\begin{aligned} \hat{I}'(l_i) &= \frac{N_1}{2} \log \frac{N_1}{N_1 + N_2} \frac{1 + l_i}{l_i} + \frac{N_2}{2} \log \frac{N_2}{N_1 + N_2} (1 + l_i) \\ &= \frac{N_1}{2} \log \frac{N_1}{N_1 + N_2} + \frac{N_2}{2} \log \frac{N_2}{N_1 + N_2} \\ &\qquad + \frac{N_1 + N_2}{2} \log (1 + l_i) - \frac{N_1}{2} \log l_i, \\ \hat{J}'(F_i) &= \frac{N_1 N_2}{2(N_1 + N_2)} \frac{(F_i - 1)^2}{F_i}. \end{aligned} \tag{7.1.11}$$

It is conjectured that for large N_1 and N_2, (assuming that the corresponding population parameters have null hypothesis values) the quantity $2\hat{I}'(l_{m+1}) + \cdots + 2\hat{I}'(l_k)$, the terms arranged in descending order of efficiency, has a chi-square distribution with $(k - m)(k - m + 1)/2$ d.f.

(7.2) *r-samples.* Suppose that we have r independent samples, respectively, of $N_1, N_2, \cdots, N_r$, independent observations each, from k-variate normal populations for which we assume the means equal, and that for the population covariance matrices we have the two hypotheses, $H_1: \Sigma_1, \Sigma_2, \cdots, \Sigma_r$ and $H_2: \Sigma_1 = \Sigma_2 = \cdots = \Sigma_r = \Sigma$. Thus, for the r samples, corresponding to H_1 and H_2, we have, respectively,

$$\sigma_{(1)} = \begin{array}{c} \begin{bmatrix} \Sigma_1 & & & & & \cdots & & 0 \\ & \ddots & & & & & & \\ & & \Sigma_1 & & & & & \\ & & & \Sigma_2 & & & & \\ \vdots & & & & \ddots & & & \vdots \\ & & & & & \Sigma_2 & & \\ & & & & & & \Sigma_r & \\ & & & & & & & \ddots \\ 0 & & & & \cdots & & & \Sigma_r \end{bmatrix} \end{array} \begin{array}{l} N_1 \\ \\ \\ \\ N_2 \\ \vdots \\ \\ \\ N_r \end{array}, \qquad N = N_1 + N_2 + \cdots + N_r, \tag{7.2.1}$$

$$\sigma_{(2)} = \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} \begin{array}{l} N_1 \\ N_2 \\ \vdots \\ N_r \end{array}, \tag{7.2.2}$$

$$\begin{aligned} I(1:2) &= \tfrac{1}{2} \log \frac{|\Sigma|^{N_1+N_2+\cdots+N_r}}{|\Sigma_1|^{N_1} \cdots |\Sigma_r|^{N_r}} + \tfrac{1}{2} \operatorname{tr} \begin{pmatrix} \Sigma_1 - \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r - \Sigma \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma^{-1} \end{pmatrix} \\ &= \sum_{i=1}^{r} \frac{N_i}{2} \left(\log \frac{|\Sigma|}{|\Sigma_i|} + \operatorname{tr} \Sigma_i \Sigma^{-1} \right) - \frac{kN}{2}, \end{aligned} \tag{7.2.3}$$

$$\begin{aligned} J(1, 2) &= \tfrac{1}{2} \operatorname{tr} \left[\begin{pmatrix} \Sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r \end{pmatrix} - \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix} \right] \\ &\quad \cdot \left[\begin{pmatrix} \Sigma^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma^{-1} \end{pmatrix} - \begin{pmatrix} \Sigma_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_r^{-1} \end{pmatrix} \right] \\ &= \sum_{i=1}^{r} \frac{N_i}{2} (\operatorname{tr} \Sigma_i \Sigma^{-1} + \operatorname{tr} \Sigma \Sigma_i^{-1}) - kN. \end{aligned} \tag{7.2.4}$$

If we compute the sample values about the sample averages, then we estimate $I(1:2)$ and $J(1,2)$, by taking $N_1, N_2, \cdots, N_r$, as degrees of freedom, and replace $\Sigma_1, \Sigma_2, \cdots, \Sigma_r$, respectively, by the sample unbiased covariance matrices $S_1, S_2, \cdots, S_r$, and Σ by S, where $NS = N_1S_1 + \cdots + N_rS_r$. Thus we have (cf. Box [5] and Wilks [31], p. 489),

$$\hat{I}(1:2) = \sum_{i=1}^{r} \frac{N_i}{2}\left(\operatorname{tr} S_i S^{-1} + \log \frac{|S|}{|S_i|}\right) - \frac{kN}{2} = \sum_{i=1}^{r} \frac{N_i}{2} \log \frac{|S|}{|S_i|}, \tag{7.2.5}$$

$$\begin{aligned} \hat{J}(1,2) &= \sum_{i=1}^{r} \frac{N_i}{2}(\operatorname{tr} S_i S^{-1} + \operatorname{tr} SS_i^{-1}) - kN = \sum_{i=1}^{r} \frac{N_i}{2} \operatorname{tr} SS_i^{-1} - \frac{kN}{2} \\ &= \sum_{i<j} \frac{N_i N_j}{2N}(\operatorname{tr} S_i S_j^{-1} + \operatorname{tr} S_j S_i^{-1} - 2k). \end{aligned} \tag{7.2.6}$$

We omit at this time a discussion of linear discriminant functions for this case.

8. Asymptotic distribution of $\hat{I}(1:2)$ for the homogeneity of covariance matrices. On the hypothesis H_2 of Section 7.2, we let

$$N_i S_i = \Sigma^{1/2} V_i \Sigma^{1/2}, \qquad NS = \Sigma^{1/2} V \Sigma^{1/2}, \qquad i = 1, 2, \cdots, r. \tag{8.1}$$

These equations define transformations linear in the elements of the matrices S_i, S or V_i, V. The Jacobians of these transformations are given by [8],

$$\left|\frac{1}{N_i}\Sigma\right|^{(k+1)/2} \quad \text{and} \quad \left|\frac{1}{N}\Sigma\right|^{(k+1)/2}.$$

The Wishart distributions of the elements of S_i, S are thereby transformed into the respective probability densities of the elements of V_i, V, given by

$$\frac{(\frac{1}{2})^{kN_i/2} e^{-(1/2)\operatorname{tr} V_i} |V_i|^{(N_i-k-1)/2}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N_i + 1 - \alpha}{2}\right)}, \qquad \frac{(\frac{1}{2})^{kN/2} e^{-(1/2)\operatorname{tr} V} |V|^{(N-k-1)/2}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N + 1 - \alpha}{2}\right)}. \tag{8.2}$$

Applying the transformations in (8.1) to $\hat{I}(1:2)$ in (7.2.5), we get

$$\hat{I}(1:2) = \sum_{\beta=1}^{r} \frac{N_\beta}{2}\left(\log \frac{|V|}{|V_\beta|} + k \log \frac{N_\beta}{N}\right). \tag{8.3}$$

Since the r samples are independent, the characteristic function of the distribution of

$$\sum_{\beta=1}^{r} N_\beta \log \frac{|V|}{|V_\beta|} = N \log |V| - \sum_{\beta=1}^{r} N_\beta \log |V_\beta|$$

is given by (cf. Box [5], p. 321)

$$\phi(t) = \int \left(\prod_{\beta=1}^{r} \frac{(\frac{1}{2})^{kN_\beta/2} e^{-(1/2)\operatorname{tr} V_\beta} |V_\beta|^{(N_\beta(1-2it)-k-1)/2}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} \right) |V|^{Nit} \prod_{\beta=1}^{r} \prod_{\gamma,\delta=1}^{k} dV_{\beta\gamma\delta}$$

$$= \left(\prod_{\beta=1}^{r} \prod_{\alpha=1}^{k} \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} \right)$$

$$(8.4) \qquad \cdot \int \frac{(\frac{1}{2})^{kN/2} e^{-(1/2)\operatorname{tr} V} |V|^{[N(1-2it)-k-1]/2+Nit} \prod_{\gamma,\delta=1}^{k} dV_{\gamma,\delta}}{\pi^{k(k-1)/4} \prod_{\alpha=1}^{k} \Gamma\left(\frac{N(1-2it)+1-\alpha}{2}\right)}$$

$$= \prod_{\alpha=1}^{k} \left(\frac{\Gamma\left(\frac{N+1-\alpha}{2}\right)}{\Gamma\left(\frac{N(1-2it)+1-\alpha}{2}\right)} \prod_{\beta=1}^{r} \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} \right),$$

where the middle result follows from the reproductive property of the Wishart distribution [33]. We will use Stirling's approximation

$$\log \Gamma(p) = \tfrac{1}{2} \log 2\pi + (p - \tfrac{1}{2}) \log p - p + \tfrac{1}{12}p - \tfrac{1}{360}p^3 + O(1/p^5)$$

to get an approximate value for large N_β in (8.4). We have that

$$\log \frac{\Gamma\left(\frac{N_\beta(1-2it)+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_\beta+1-\alpha}{2}\right)} =$$

$$(8.5) \qquad \begin{aligned} &\frac{N_\beta(1-2it)-\alpha}{2} \cdot \log \frac{N_\beta(1-2it)+1-\alpha}{2} - \frac{N_\beta(1-2it)+1-\alpha}{2} \\ &+ \frac{1}{6(N_\beta(1-2it)+1-\alpha)} - \frac{1}{45(N_\beta(1-2it)+1-\alpha)^3} \\ &- \frac{N_\beta-\alpha}{2} \log \frac{N_\beta+1-\alpha}{2} \\ &+ \frac{N_\beta+1-\alpha}{2} - \frac{1}{6(N_\beta+1-\alpha)} + \frac{1}{45(N_\beta+1-\alpha)^3} + O(1/N_\beta^5), \end{aligned}$$

and after some algebraic manipulation, the right member of (8.5) may be written as

$$-itN_\beta \log \frac{N_\beta}{2} + \frac{N_\beta(1-2it)-\alpha}{2} \log (1-2it) + N_\beta it + \frac{(3\alpha^2-1)2it}{12N_\beta(1-2it)} + O(1/N_\beta^2).$$

We therefore have

$$
\begin{aligned}
\log \phi(t) = {} & \sum_{\alpha=1}^{k}\left(itN \log \frac{N}{2} - \frac{N(1-2it)-\alpha}{2}\cdot \log(1-2it)\right.\\
& \left. - Nit - \frac{(3\alpha^2-1)it}{6N(1-2it)} - 0(1/N^2)\right)\\
& + \sum_{\alpha=1}^{k}\sum_{\beta=1}^{r} - itN_\beta \log \frac{N_\beta}{2} + \frac{N_\beta(1-2it)-\alpha}{2}\cdot \log(1-2it)\\
& + N_\beta it + \frac{(3\alpha^2-1)it}{6N_\beta(1-2it)} + 0(1/N_\beta^2)\\
= {} & -it\sum_{\beta=1}^{r} kN_\beta \log \frac{N_\beta}{N} - \frac{(r-1)k(k+1)}{4}\log(1-2it)\\
& + \frac{it(2k^3+3k^2-k)}{12(1-2it)}\left(\sum_{\beta=1}^{r}\frac{1}{N_\beta} - \frac{1}{N}\right) + \sum_{\beta=1}^{r} 0(1/N_\beta^2) - 0(1/N^2).
\end{aligned}
\tag{8.6}
$$

Neglecting the last term in (8.6), we have that

$$
\phi(t) = (1-2it)^{-(r-1)k(k+1)/4}\exp\left(-it\sum_{\beta=1}^{r} kN_\beta \log \frac{N_\beta}{N} + \frac{Cit}{1-2it}\right), \tag{8.7}
$$

where $C = (2k^3 + 3k^2 - k)\left(\sum_{\beta=1}^{r} 1/N_\beta - 1/N\right)/12$.

Because of (8.3) and (8.4), writing $\zeta = 2\hat{I}(1:2)$, the probability density of ζ is given by

$$
D(\zeta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-it\zeta + Cit/(1-2it)}\,dt}{(1-2it)^{(r-1)k(k+1)/4}}. \tag{8.8}
$$

If we neglect the term with C, it follows that $D(\zeta)$ is a chi-square distribution with $(r-1)k(k+1)/2$ d.f.; otherwise, by integrating (8.8) (see [23], p. 86), we get, since ζ is real and positive and $(r-1)k(k+1)/4 > 0$,

$$
D(\zeta) = \tfrac{1}{2}e^{-C/2-\zeta/2}\left(\frac{\zeta}{C}\right)^{(n-1)/2} I_{n-1}(\sqrt{C\zeta}), \tag{8.9}
$$

where $n = (r-1)k(k+1)/4$ and $I_{n-1}(\sqrt{C\zeta})$ is the Bessel function of purely imaginary argument [30]

$$
I_{n-1}(\sqrt{C\zeta}) = \sum_{j=0}^{\infty} \frac{(\frac{1}{2})^{(n-1)/2+j}\left(\dfrac{C\zeta}{2}\right)^{(n-1)/2+j}}{j!\Gamma(n+j)}.
$$

The distribution given by (8.9) is the non-central chi-square distribution and is Fisher's B distribution ([10, p. 14. 665) if we write $C = \beta^2$, $\zeta = B^2$, $2n = n_1$. The case for $k = 1$ is the Bartlett test for homogeneity of variance [3], [5].

The approximation to the logarithm of the characteristic function of ζ, i.e., $-n \log (1 - 2it) + cit/(1 - 2it)$, corresponds to that of Box [5], formula 29,

p. 323, retaining only the first term in his sum; i.e., $(\alpha_1/\mu)[1/(1 - 2it) - 1]$ (there is a misprint in the formula) is $Cit/(1 - 2it)$ as used here, as may be verified by using the appropriate formulas with $\beta = 0$ on pp. 324–325 of [5].

For large n we may approximate $I_{n-1}(\sqrt{C\zeta})$ in (8.9) by writing

$$\begin{aligned} I_{n-1}(\sqrt{C\zeta}) &= \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{(C\zeta/4)^j \Gamma(n)}{j!\Gamma(n+j)} \\ &\approx \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{C\zeta}{4n}\right)^j \\ &= \frac{(C\zeta/4)^{(n-1)/2}}{\Gamma(n)} e^{C\zeta/4n} \end{aligned}$$

and thereby get

$$D(\zeta) \approx \tfrac{1}{2} \frac{e^{-C/2-\zeta[1-(C/2n)]/2}}{\Gamma(n)} \left(\frac{\zeta}{2}\right)^{n-1}. \tag{8.10}$$

If we set $\zeta[1 - (C/2n)] = \chi^2$, (8.10) yields

$$\begin{aligned} D(\chi^2)\, d\chi^2 &= \frac{e^{-C/2}}{\left(1 - \dfrac{C}{2n}\right)^n} \cdot \frac{e^{-\chi^2/2}}{\Gamma(n)} \left(\frac{\chi^2}{2}\right)^{n-1} d\,\frac{\chi^2}{2} \\ &\approx \frac{e^{-\chi^2/2}(\chi^2/2)^{n-1}\, d\chi^2/2}{\Gamma(n)}, \end{aligned} \tag{8.11}$$

or $\zeta[1 - (C/2n)]$ asymptotically has a chi-square distribution with

$$2n = (r - 1)k(k + 1)/2 \text{ d.f.}$$

It is readily verified that $1 - (C/2n) = \rho$, Box's scale factor in the chi-square approximation ([5], p. 329).

For other approximations to (8.9), see Abdel-Aty [1].

Examples: (a). For the first example we use the data given by Smith [29], Table 2, which he used to calculate a linear discriminant function for a group of 25 normal persons and 25 psychotics. Here $k = 2$, $r = 2$,

$$S_1 = \begin{pmatrix} 6.92 & -5.27 \\ -5.27 & 40.89 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 36.75 & 13.92 \\ 13.92 & 287.92 \end{pmatrix}, \quad S = \begin{pmatrix} 21.83 & 4.33 \\ 4.33 & 164.40 \end{pmatrix},$$

$$N_1 = N_2 = 24, \quad N = 48, \quad |S_1| = 255.1859, \quad |S_2| = 10387.2936,$$

$$|S| = 3570.1031,$$

$$2\hat{I} = 24 \log (3570.1031/255.1859) + 24 \log (3570.1031/10387.2936) = 37.7268,$$

$$C = (16 + 12 - 2)(2/24 - 1/48)/12 = .135416 = \beta^2, \quad \beta = .368,$$

$$n = (2 - 1)(2)(3)/4, \quad 2n = 3 = n_1,$$

$$\zeta = 2\hat{I} = 37.7268 = B^2, \quad B = 6.14.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 3$ and $\beta = .2$ and .4 to be, respectively, 2.8140 and 2.8680. We therefore reject the null hypothesis of equality of the population covariance matrices. Smith [29] does remark that the correlations are not significant, but the variances of the psychotics are significantly greater than those of the normals.

(b) For the second example, we use the data given by Kossack [19] for a problem of classifying an A.S.T.P. pre-engineering trainee as to whether he would do unsatisfactory or satisfactory work in his first-term mathematics course. The three variables used are x_1, a mathematics placement test score; x_2, a high school mathematics score; x_3, the Army General Classification Test score. There were 96 trainees who did unsatisfactory work and 209 who performed satisfactory work. Here $k = 3$, $r = 2$, $N_1 = 95$, $N_2 = 208$, $N = 303$,

$$S_1 = \begin{pmatrix} 133.8592 & 7.0572 & 2.0717 \\ 7.0572 & 4.1288 & -2.0109 \\ 2.0717 & -2.0109 & 27.7016 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 217.1505 & 14.0692 & 35.7085 \\ 14.0692 & 3.9820 & .4031 \\ 35.7085 & .4031 & 72.7206 \end{pmatrix},$$

$$S = \begin{pmatrix} 191.04 & 11.871 & 25.162 \\ 11.871 & 4.0280 & -.35378 \\ 25.162 & -.35378 & 58.606 \end{pmatrix}, \quad |S_1| = 13313, \quad |S_2| = 43779,$$

$$|S| = 34053, 2\hat{I} = 95 \log \frac{34053}{13313} + 208 \log \frac{34053}{43779} = 227.0867,$$

$$C = (54 + 27 - 3)(1/95 + 1/208 - 1/303)/12 = .078221 = \beta^2, \quad \beta = .28,$$

$$n = (2 - 1)(3)(4)/4, \quad 2n = 6 = n_1,$$

$$\zeta = 2\hat{I} = 227.0867 = B^2, \quad B = 15.06.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 6$ and $\beta = .2$ and .4 to be, respectively, 3.5602 and 3.5951. We therefore reject the null hypothesis of equality of the population covariance matrices. An assumption of equality is, however, implicit in the procedure used by Kossack.

(c) For the third example, we use the data given by Pearson and Wilks [24], for five samples of twelve observations each on the strength and hardness in aluninum die-castings. Based on their data (note that they did not use the unbiased estimates), the details of which are not repeated here,

$$k = 2, \quad r = 5, \quad N_1 = \cdots = N_5 = 11, \quad N = 55,$$

$$\log |S_1| = 5.82588, \quad \log |S_2| = 6.63942, \quad \log |S_3| = 5.31904,$$

$$\log |S_4| = 6.66973, \quad \log |S_5| = 5.35937, \quad \log |S| = 6.13953,$$

$$2\hat{I} = 55(6.13953) - 11(29.81344) = 9.726,$$

$$C = (16 + 12 - 2)(5/11 - 1/55)/12 = .945454 = \beta^2, \quad \beta = .972,$$

$$n = (5 - 1)(2)(3)/4, \quad 2n = 12 = n_1,$$

$$\zeta = 2\hat{I} = 9.726 = B^2, \quad B = 3.12.$$

In Fisher's B Table ([10], p. 14.665) we find the 5 per cent points for $n_1 = 7$ (the largest there tabulated), and $\beta = 0.8$ and 1.0 to be, respectively, 3.9144 and 4.0005. Since the tabulated values increase with increasing n_1 for a fixed β, we do not, in this case, reject the null hypothesis of equality of population covariance matrices. This is consistent with the conclusion reached by Pearson and Wilks [24].

9. Asymptotic distribution of $\hat{I}'_R$. In (4.9) we defined $\hat{I}'_R$ and made certain statements about its asymptotic distribution which we will now confirm.

It is known that the logarithm of the characteristic function of the distribution of $2\hat{I}'_R$ is given by (see [31], p. 492; [4])

$$\log \phi(t) = (k-1)\log \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N}{2} - Nit\right)} + \sum_{\alpha=1}^{k-1} \log \frac{\Gamma\left(\frac{N(1-2it)-\alpha}{2}\right)}{\Gamma\left(\frac{N-\alpha}{2}\right)}. \tag{9.1}$$

Employing Stirling's approximation as in (8.5), and retaining comparable terms as in (8.7), we have

$$\log \phi(t) = -\frac{k(k-1)}{4}\log(1-2it) + \frac{Cit}{1-2it}, \tag{9.2}$$

where $C = k(k-1)(2k+5)/12N$.

The statement at the end of Section 4 then follows from (9.2), (8.8), and (8.9). From (8.11) we may also deduce that

$$2\hat{I}'_R\left(1 - \frac{k(k-1)(2k+5)}{6Nk(k-1)}\right) = -(N - \tfrac{1}{6}(2k+5))\log|R|$$

asymptotically has a chi-square distribution with $k(k-1)/2$ d.f. This latter result is given by Bartlett [4].

10. Asymptotic distribution of $\hat{J}(1,2)$ for the linear hypothesis. From results derived by Fisher [9], Girshick [11], Hsu [15], [16], [17], and Roy [27], it is known that the probability density of the distribution of the roots of $|S^* - lS| = 0$ (see (5.18)), for $(n-r)$ large, is given by

$$\frac{(\frac{1}{2})^{(r-1)p/2}\pi^{p/2}(l_1 \cdots l_p)^{(r-p-2)/2}}{\prod_{\alpha=1}^{p}\Gamma\left(\frac{r-\alpha}{2}\right)\Gamma\left(\frac{p+1-\alpha}{2}\right)} e^{-\frac{1}{2}(l_1+\cdots+l_p)} \prod_{i>j}(l_j - l_i) \tag{10.1}$$

and that of the roots of $|S_{21}S_{11}^{-1}S_{12} - lS_{22\cdot1}| = 0$ (see (6.3), (6.4)), for $(n-k_1)$ large, is given by

$$\frac{(\frac{1}{2})^{k_1k_2/2}\pi^{k_2/2}(V_1 \cdots V_{k_2})^{(k_1-k_2-1)/2}}{\prod_{\alpha=1}^{k_2}\Gamma\left(\frac{k_1+1-\alpha}{2}\right)\Gamma\left(\frac{k_2+1-\alpha}{2}\right)} e^{-\frac{1}{2}(V_1+\cdots+V_{k_2})} \prod_{i>j}(V_j - V_i), \tag{10.2}$$

where $V_i = (n-k_1)l_i$.

From (10.1) and (10.2) it is readily derived that the characteristic functions of the asymptotic distributions of $2\hat{I}(1:2) = \hat{J}(1,2)$ in (5.18) and (6.4) are, respectively, $(1 - 2it)^{-(r-1)p/2}$ and $(1 - 2it)^{-k_1k_2/2}$, whence the conclusion as to their chi-square distributions. The chi-square decompositions in Section 5 and Section 6 follow from the fact that asymptotically the distributions of

$$l_{m+1}, \cdots, l_p$$

of (10.1) and $V_{m+1}, \cdots, V_{k_2}$ of (10.2), assuming that the corresponding population parameters have the null hypothesis values, are independent of the distribution of the remaining roots and are given, respectively, by

$$\frac{(\tfrac{1}{2})^{(r-1-m)(p-m)/2}\pi^{(p-m)/2}}{\prod_{\alpha=m+1}^{p}\Gamma\left(\frac{r-m-\alpha}{2}\right)\Gamma\left(\frac{p-m+1-\alpha}{2}\right)} (l_{m+1}\cdots l_p)^{(r-p-2)/2}e^{-\frac{1}{2}(l_{m+1}+\cdots+l_p)}\prod_{i>j}(l_j - l_i), \tag{10.3}$$

$$\frac{(\tfrac{1}{2})^{(k_1-m)(k_2-m)/2}\pi^{(k_2-m)/2}}{\prod_{\alpha=m+1}^{k_2}\Gamma\left(\frac{k_1-m+1-\alpha}{2}\right)\Gamma\left(\frac{k_2-m+1-\alpha}{2}\right)} (V_{m+1}\cdots V_{k_2})^{(k_1-k_2-1)/2}e^{-\frac{1}{2}(V_{m+1}+\cdots+V_{k_2})}\prod_{i>j}(V_j - V_i). \tag{10.4}$$

The characteristic function of the distribution of $2\hat{I}$ of (7.1.10) could also have been derived from the distribution of the roots of $|N_1S_1 - lN_2S_2| = 0$, given by,

$$\pi^{k/2}\left(\prod_{\alpha=1}^{k}\frac{\Gamma\left(\frac{N_1+N_2+1-\alpha}{2}\right)}{\Gamma\left(\frac{N_1+1-\alpha}{2}\right)\Gamma\left(\frac{N_2+1-\alpha}{2}\right)\Gamma\left(\frac{k+1-\alpha}{2}\right)}\right) \frac{(l_1\cdots l_k)^{(N_1-k-1)/2}\prod_{i>j}(l_j - l_i)}{((1+l_1)\cdots(1+l_k))^{(N_1+N_2)/2}}. \tag{10.5}$$

11. Concluding remarks. The validity of the conjecture at the end of Section 7.1 is under investigation, as well as the distributions of $\hat{J}$ and $\hat{J}'(F_i)$ of Section 7, and related power functions.

It might also be mentioned that we have a basis for assessing the cost of trading observations for dimensions. If there is more than one significant linear discriminant function, then N_1 observations with the linear function associated with λ_1 (one dimension) would be as effective as N observations with the original multidimensional variables, where $NJ(1,2) = N_1J'(1,2;\lambda_1)$. Similar conclusions hold for more than one linear function.

Procedures similar to those used herein to estimate $I(1:2)$ and $J(1,2)$ are also applicable to problems of testing appropriate hypotheses for other than normal populations.

I am indebted to the referees for comments which have contributed to improvements in this paper.

REFERENCES

[1] S. H. Abdel-Aty, "Approximate formulae for the percentage points and the probability integral of the non-central X^2 distribution," *Biometrika*, Vol. 41 (1954), pp. 538–540.
[2] T. W. Anderson, "The asymptotic distribution of certain characteristic roots and vectors," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 103–130.
[3] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc. London*, Ser. A, Vol. 160 (1937), pp. 268–282.
[4] M. S. Bartlett, "Tests of significance in factor analysis," *Brit. J. Psychology, Stat. Sec.*, Vol. 3 (1950), pp. 77–85.
[5] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, Vol. 36 (1949), pp. 317–346.
[6] H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilitiés," *Actualités Scientifiques et Industrielles*, No. 736, Hermann & Cie, Paris, 1938.
[7] H. E. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 631–650.
[8] W. L. Deemer and I. Olkin, "The Jacobians of certain matrix transformations useful in multivariate analysis," *Biometrika*, Vol. 38 (1951), pp. 345–367.
[9] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.
[10] R. A. Fisher, *Contributions to Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1950.
[11] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203–224.
[12] S. W. Greenhouse, "On the problem of discrimination between statistical populations," M. A. Thesis, The George Washington University, 1954.
[13] H. Hotelling, "A generalized T test and measure of multivariate dispersion," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 23–41.
[14] J. P. Hoyt, "Estimates and asymptotic distributions of certain statistics in information theory," Dissertation, The Graduate Council of The George Washington University, 1953.
[15] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.
[16] P. L. Hsu, "On the limiting distribution of the canonical correlations," *Biometrika*, Vol. 32 (1941–42), pp. 38–45.
[17] P. L. Hsu, "On the limiting distribution of roots of a determinantal equation," *J. London Math. Soc.*, Vol. 16 (1941), pp. 183–194.
[18] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, New York, 1949.
[19] C. F. Kossack, "On the mechanics of classification," *Ann. Math. Stat.*, Vol. 16 (1945) pp. 95–98.
[20] S. Kullback, "An application of information theory to multivariate analysis," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 88–102.
[21] S. Kullback, "Certain inequalities in information theory and the Cramér-Rao inequality," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 745–751.
[22] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.* Vol. 22 (1951), pp. 79–86.

[23] N. W. McLachlan, *Complex Variable and Operational Calculus with Technical Applications*, Cambridge University Press, 1939.
[24] E. S. Pearson and S. S. Wilks, "Methods of statistical analysis appropriate for k samples of two variables," *Biometrika*, Vol. 25 (1933), pp. 353–378.
[25] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, Inc., New York, 1952.
[26] D. D. Rippe, "Statistical rank and sampling variation of the results of factorization of covariance matrices," Doctoral Thesis on file at the University of Michigan, 1951.
[27] S. N. Roy, "p-statistics, or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.
[28] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220–238.
[29] C. A. B. Smith, "Some examples of discrimination," *Ann. Eugenics*, Vol. 13 (1947), pp. 272–282.
[30] G. N. Watson, *Bessel Functions*, (2d ed.), The Macmillan Co., New York, 1944.
[31] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–491.
[32] S. S. Wilks, "The large sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 60–62.
[33] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1943.

ON THE CHARACTERISTICS OF THE GENERAL QUEUEING PROCESS, WITH APPLICATIONS TO RANDOM WALK[1]

BY J. KIEFER AND J. WOLFOWITZ

Cornell University

Summary. The authors continue the study (initiated in [1]) of the general queueing process (arbitrary distributions of service time and time between successive arrivals, many servers) for the case ($\rho < 1$) where a limiting distribution exists. They discuss convergence with probability one of the mean waiting time, mean queue length, mean busy period, etc. Necessary and sufficient conditions for the finiteness of various moments are given. These results have consequences for the theory of random walk, some of which are pointed out.

This paper is self-contained and may be read independently of [1]; the necessary results of [1] are quoted. No previous knowledge of the theory of queues is required for reading either [1] or the present paper.

Introduction. We recapitulate very briefly some of the results obtained in [1] in the notation of [1] to which we shall adhere without further mention.[2]

Let S be the totality of points $(x_1, x_2, \cdots, x_s)$ of Euclidean s-space such that $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_s$. Let x and y be generic points of S. Occasionally another letter will represent a point in S; it will always be clear from the context when this is so; for example, O will frequently denote the origin in s-space.

For $i \geqq 1$, let $t_i \geqq t_0 = 0$ be the time of arrival of the ith person at a system of $s \geqq 1$ machines, where he waits his turn until a machine is available to serve him, say at time $t_i + w_{i1} \geqq t_i$. This machine is then occupied by him for time $R_i \geqq 0$. Let $g_i = t_i - t_{i-1}$. $\{R_i\}$ and $\{g_i\}$ are independent sequences of identically distributed and independent chance variables. An s-dimensional random walk $\{w_i\}$, with w_{i1} its first component, is useful for the study of the theory of queues. The random walk $\{w_i\}$ is constructed as follows: $w_i = (w_{i1}, \cdots, w_{is})$. Unless the contrary is explicitly stated we have $w_1 = O$. To obtain w_{i+1} from w_i, reorder in ascending size the quantities

$$(w_{i1} + R_i - g_{i+1})^+, \qquad (w_{i2} - g_{i+1})^+, \qquad (w_{i3} - g_{i+1})^+, \cdots, (w_{is} - g_{i+1})^+.$$

The resulting sequence is w_{i+1}. We have $w_{i1} \leqq w_{i2} \leqq \cdots \leqq w_{is}$ for all i. As usual, $a^+ = (a + |a|)/2$. The times $t_i + w_{ij}$ $(1 \leqq j \leqq s)$ are easily seen to be the earliest times after (or at) t_i at which the s machines have finished serving those of the first $s - 1$ arrivals which they serve.

Let $F_i(F_i^*)$ be the d.f. (distribution function) of $w_i(w_{i1})$. It was shown in [1] that $F(x) = \lim_{i\to\infty} F_i(x)$ exists and satisfies a certain integral equation (I.E.);

Received Dec. 21, 1954.

[1] Research under contract with the Office of Naval Research.

[2] The definition of ν on p. 14 of [1] should be modified trivially to read $\nu = 1$ in the case $b = \infty$.

$F^*(z) = \lim_{i\to\infty} F_i^*(z)$ also exists. Assume $\rho = ER_i / sEg_i$ exists. F and F^* are d.f.'s if $\rho < 1$, and F is then the unique d.f. solution to the I.E. Except in the trivial case where $P\{R_i = sg_i\} = 1$, if $\rho \geqq 1$ then $F \equiv 0 \equiv F^*$, and the I.E. has no d.f. solution. Always $F^*(z) = F(z, \infty, \cdots, \infty)$. Results on the limiting length of the line are also proved in [1].

Let $F_i(x \mid y)$ be the d.f. of w_i, given that $w_1 = y$; i.e.,

$$F_i(x \mid y) = P\{w_i \leqq x \mid w_1 = y\}.$$

It was proved in [1] that, for all $y \varepsilon S$,

$$\lim_{i\to\infty} F_i(x \mid y) = F(x).$$

Throughout this paper we shall assume that $\rho < 1$. The case $\rho \geqq 1$ has little interest and was essentially disposed of in [1]; results proved in the present paper are trivial when $\rho \geqq 1$. Throughout this paper we shall assume that $Eg_1 < \infty$. However, it can be shown, always easily and sometimes trivially, that all the results of [1] and all the queueing results of the present paper except Theorem 3 are valid also when $Eg_1 = \infty$. In order to eliminate the completely trivial we also assume, as was done in [1], that $ER_1 > 0$, $Eg_1 > 0$. Since $\rho < 1$ we have then $0 < ER_1 < \infty$, $0 < Eg_1 < \infty$.

In two or three places below we shall cite the first paragraph of Section 3 of [1]. To ease the reader's task we now quote this paragraph in full:

Let $\varphi_j(a, b, c)$, $j = 1, \cdots, s$ be the value of $w_{(i+1),j}$ when $w_i = a$, $R_i = b$, $g_{i+1} = c$. If d is a point in s-space, we shall say that $a \leqq d$ if every coordinate of a is not greater than the corresponding coordinate of d. If now $a \leqq d$, then obviously

$$\varphi_j(a, b, c) \leqq \varphi_j(d, b, c)$$

for $1 \leqq j \leqq s$. Applying this argument k times we obtain the following result: Let $R_{i+j-1} = b_{i+j-1}$, $g_{i+j} = c_{i+j}$, $j = 1, \cdots, k$. Let $w_{i+k} = e_1$ when $w_i = a_1$, and let $w_{i+k} = e_2$ when $w_i = a_2$. Then $a_1 \leqq a_2$ implies $e_1 \leqq e_2$.

The results of [1] also imply that $F(x)$ determines a stationary and metrically transitive flow; this is the process $\{w_n^0\}$ defined in Section 1, below, where the relevant references to [1] are given.

1. Convergence of the mean waiting time. Let k be any positive number. Define $W_n = \sum_{i=1}^{s} w_{ni}$. Since w_{ni} is a nonnegative chance variable and $F_n(x) \to F(x)$, we easily have that

$$(1.1)\qquad \begin{aligned} \liminf_n (Ew_{ni})^k &\geqq \int (x_i)^k \, dF(x), \\ \liminf_n E(W_n)^k &\geqq \int (x_1 + \cdots + x_s)^k \, dF(x), \end{aligned}$$

where, of course, the right members may be infinite. From the fact (proved in

[1]) that $F_n(x)$ approaches $F(x)$ from above for every x, we have that

$$E(w_{ni})^k \leqq \int (x_i)^k \, dF(x).$$

Hence

$$\lim_n E(w_{ni})^k = \int (x_i)^k \, dF(x). \tag{1.2}$$

Let $F_n^W(z \mid y)$ be the d.f. of W_n, given that $w_1 = y(\varepsilon S)$. Hence $F_n^W(z \mid 0)$ is the d.f. of W_n. Then

$$F_{n+1}^W(z \mid 0) - F_n^W(z \mid 0) = \int [F_n^W(z \mid y) - F_n^W(z \mid 0)] \, dF_2(y).$$

It follows from the first paragraph of Section 3 of [1] that, if $y \varepsilon S$, the integrand in the last integral is never positive for any z. Hence the left member in the last equation is never positive for any z. Hence $F_n^W(z \mid 0)$ approaches its limit (which is a distribution function obtainable from $F(x)$ in an obvious way) from above. Consequently, as before,

$$E(W_n)^k \leqq \int \left(\sum_{i=1}^{s} x_i\right)^k dF(x).$$

From this and (1.1) we obtain

$$\lim_n E(W_n)^k = \int \left(\sum_{i=1}^{s} x_i\right)^k dF(x) = m_k' \text{ (say)}.$$

The question as to when $m_k' < \infty$ will be discussed in a later section. We define

$$m_k = \int (x_1)^k \, dF(x),$$

and

$$V_{nk} = \frac{1}{n} \sum_{i=1}^{n} (w_{i1})^k.$$

We now prove

THEOREM 1. *We have, for any positive k,*

$$P\{\lim_{n \to \infty} V_{nk} = m_k\} = 1. \tag{1.3}$$

PROOF. Let w_1^0 be an s-dimensional chance variable with the d.f. $F(x)$, and let w_{n+1}^0 be obtained from w_n^0 by using R_n and g_{n+1} in exactly the same manner as one obtains w_{n+1} from w_n. Thus w_n^0 pertains at time t_n. Then the process $\{w_n^0, n = 1, 2, \cdots\}$ is easily seen to be stationary, because $F(x)$ satisfies the integral equation derived in [1] (see Section 3 of [1] for details). It is proved in Section 8 of [1] that $F(x)$ is the only d.f. which satisfies the integral equation.

We shall show that this implies easily that there cannot be a Borel set B in s-dimensional Euclidean space such that

$$0 < \int_B dF < 1,$$

and $w_1^0 \varepsilon B$ implies with probability one that $w_n^0 \varepsilon B$, $n \geqq 2$. For let $\bar{B}$ be the complement of B, and $F(x \mid B)$ and $F(x \mid \bar{B})$ be, respectively, the conditional distribution functions on B and $\bar{B}$ implied by $F(x)$. Then $F(x \mid B)$ satisfies the integral equation. On a set of w_1^0 of probability one according to $F(x \mid \bar{B})$, $w_n^0 \varepsilon \bar{B}$ for $n \geqq 2$ with probability one, since otherwise $P\{w_n^0 \varepsilon B\}$ (when F is the distribution function of w_1^0) would not be independent of n, contradicting the stationarity of $\{w_n^0\}$. Hence $F(x \mid \bar{B})$ must also satisfy the integral equation. Clearly, $F(x \mid B)$ and $F(x \mid \bar{B})$ are not identical, in contradiction to the fact that $F(x)$ is the only d.f. that satisfies the integral equation. From the fact that there is no invariant set B such that $0 < \int_B dF < 1$, the fact that w_n^0 is a Markoff process, and Theorem 1.1, page 460 of [6] (which asserts that any set in the space of the Markoffian chance variables $w_1^0, w_2^0, \cdots$ that is invariant under a shift transformation differs from a set B by a set of probability zero), we conclude that the process w_n^0 is metrically transitive. Hence, by the ergodic theorem,

$$P\{\lim_{n\to\infty} V_{nk}^0 = m_k\} = 1, \tag{1.4}$$

where

$$V_{nk}^0 = \frac{1}{n} \sum_{i=1}^{n} (w_{i1}^0)^k,$$

and of course w_{i1}^0 is the first component of the vector w_i^0.

From the argument in the first paragraph of Section 3 of [1], it follows that always

$$V_{nk} \leqq V_{nk}^0. \tag{1.5}$$

Hence

$$P\{\limsup_{n\to\infty} V_{nk} \leqq m_k\} = 1. \tag{1.6}$$

We shall prove that also

$$P\{\liminf_{n\to\infty} V_{nk} \geqq m_k\} = 1. \tag{1.7}$$

This will prove the theorem.

We shall now deduce (1.7) from (1.4), and for this purpose divide the argument into consideration of the four cases of Section 8 of [1]. As there defined, denote by $[a, b]$ and $[c, d]$ the smallest closed intervals for which

$$P\{a \leqq R_1 \leqq b\} = P\{c \leqq g_1 \leqq d\} = 1.$$

Of course, b or d or both may be $+\infty$.

CASE 1: $b > sc$. Let t be so large that the point $T = (t, t, \cdots, t)$ of S is such that

$$\int_{x<T} dF(x) > 0.$$

It follows from (1.4) that there exists in S a point $x < T$ such that

$$P\{\lim_{n\to\infty} V^0_{nk} = m_k \mid w^0_1 = x\} = 1. \tag{1.8}$$

It is proved in [1] that there exists an integer r such that $P\{w_{(sr)} > T\} > 0$, say $=\alpha$. From this it follows that

$$P\{w_n > T \quad \text{for at least one } n\} = 1. \tag{1.9}$$

Let h be the smallest index n for which $w_n > T$; $h < \infty$ with probability one. Obviously $R_h, R_{h+1}, \cdots$ and $g_{h+1}, g_{h+2}, \cdots$ are distributed independently of h and w_h, and have the same distribution as $R_1, R_2, \cdots$ and $g_2, g_3, \cdots$. Consequently, if we define, for $n > h$,

$$V_{nk}(h) = \frac{(w_{h1})^k + (w_{(h+1),1})^k + \cdots + (w_{n,1})^k}{n},$$

we have, using (1.8) and the argument in the first paragraph of Section 3 of [1], that

$$P\{\liminf_{n\to\infty} V_{nk}(h) \geqq m_k\} = 1. \tag{1.10}$$

Obviously from the definition of $V_{nk}(h)$ it follows that

$$P\{\lim_{n\to\infty} (V_{nk}(h) - V_{nk}) = 0\} = 1. \tag{1.11}$$

The desired result (1.7) follows from (1.10) and (1.11).

CASE 2: $a < d$. It is proved in Section 8 of [1] that, in this case,

$$P\{w^0_n = 0 \quad \text{for some } n \geqq 1\} = 1. \tag{1.12}$$

The desired result (1.3) follows from (1.4) and (1.12) by means of an argument like that in Case 1.

CASE 3: $c = d \leqq a = b < sc$. It is proved in [1] that in this case there is a point in S, there called $\bar{w}$, such that

$$P\{w_n = w_{n+1} = \cdots = \bar{w} \quad \text{for some } n \geqq 1\} = 1. \tag{1.13}$$

The desired result (1.3) follows at once.

CASE 4: $d \leqq a$, $b \leqq sc$, *and either* $a < b$ *or* $c < d$. It is proved in [1] that, in this case, there exists an $\epsilon > 0$ such that the set

$$\Gamma^\epsilon = \{y \mid y \in S, y \leqq \bar{w}^\epsilon\},$$

where

$$\bar{w}^\epsilon = (0, u^\epsilon_{s-1}, u^\epsilon_{s-2}, \cdots, u^\epsilon_1)$$

and

$$u_j^\epsilon = \max\,(0, b - jc - \epsilon),$$

has the following properties:

(a) $$P\{w_n^0 \,\varepsilon\, \Gamma^\epsilon \quad \text{for some } n \geqq 1\} = 1.$$

(This implies at once that

$$\int_{\Gamma^\epsilon} dF(x) > 0.) \tag{1.14}$$

(b) $$P\{w_s > \bar{w}^\epsilon\} > 0.$$

(This implies, using the argument in the first paragraph of Section 3 of [1], that

$$P\{w_n > \bar{w}^\epsilon \quad \text{for at least one } n > 1\} = 1.) \tag{1.15}$$

The desired result now follows exactly as in Case 1, the place of T being taken by $\bar{w}^\epsilon$.

In exactly the same manner as that employed in this section we could have proved that

$$P\left\{\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (W_i)^k = m_k'\right\} = 1 \tag{1.16}$$

and similar theorems about other moments.

2. Generalization of the lemma of Section 4 of [1]. We shall prove the following essential generalization of the fundamental lemma of Section 4 of [1] both for its use as a tool in a subsequent section and for its intrinsic interest:

LEMMA. *If, for any positive* $k > 0$,

$$ER_1^{k+1} < \infty, \tag{2.1}$$

then

$$\sup_n E(w_{ns} - w_{n1})^k < \infty; \tag{2.2}$$

or, what is equivalent,

$$\sup_n E\left((s-1)w_{ns} - \sum_{j=1}^{s-1} w_{nj}\right)^k < \infty. \tag{2.3}$$

PROOF. Define Y_i exactly as in (4.5) of [1], i.e.,

$$Y_i = \max[(s-1)R_i,\ (s-1)R_{i-1} - R_i,\ (s-1)R_{i-2} - R_{i-1} - R_i,\ \cdots,\ (s-1)R_1 - R_2 - \cdots - R_i].$$

Then (4.6) of [1] is

$$L(y', n) = P\{Y_n \leqq y'\} = P\{R_1 \leqq hy',\ R_2 \leqq h(R_1 + y'),\ \cdots,\ R_n \leqq h(R_1 + \cdots + R_{n-1} + y')\}, \tag{2.4}$$

where $h = (s-1)^{-1}$. Let $H(z)$ be the d.f. of R_1.

Define $L(y', 0) = 1$. Obviously $L(y', n)$ is nonincreasing in n and, for $n \geqq 0$,

$$\begin{aligned} L(y', n) - L(y', n+1) &= P\{Y_n \leqq y', R_{n+1} > h(R_1 + \cdots + R_n + y')\} \\ &\leqq P\{R_{n+1} > h(R_1 + \cdots + R_n + y')\} \qquad (2.5) \\ &\leqq E\{1 - H(h[R_1 + \cdots + R_n + y'])\}. \end{aligned}$$

Hence

$$\begin{aligned} 1 - L(y', n) &= \sum_{i=1}^{n} [(Ly', i-1) - L(y', i)] \qquad (2.6) \\ &\leqq \sum_{i=0}^{\infty} E\{1 - H(h[R_1 + \cdots + R_i + y'])\}. \end{aligned}$$

Let d be a small positive number and define

$$D_i = d \text{ when } R_i \geqq \frac{d}{h}$$

$$D_i = 0 \text{ otherwise.}$$

We choose d so small that $d < 1$ and

$$p = P\{D_1 = d\} > 0.$$

(We have earlier excluded the trivial case where $R_i = 0$ with probability one.) Since $R_i \geqq D_i/h$, if we replace the former by the latter in the right member of (2.6) we do not diminish any term of this member. It is well known (e.g., [2], p. 101) from approximations to the binomial distribution that, for suitable positive c_1, c_2, we have

$$P\left\{D_1 + \cdots + D_n \leqq \frac{npd}{2}\right\} < c_1 e^{-c_2 n} \qquad (2.7)$$

When $k \geqq 1$ we have, from (2.6),

$$\begin{aligned} E(Y_n)^k &\leqq k \sum_{j=0}^{\infty} (j+1)^{k-1} P\{Y_n > j\} \\ &\leqq k \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} (j+1)^{k-1} E\{1 - H(h[R_1 + \cdots R_i + j])\} \\ &\leqq k \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} (j+1)^{k-1} E\{1 - H(D_1 + \cdots D_i + j)\} \qquad (2.8) \\ &\leqq k \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} (j+1)^{k-1} E\{1 - H(D_1 + \cdots + D_{i+j})\} \\ &\leqq \sum_{j=0}^{\infty} (j+2)^k E\{1 - H(D_1 + \cdots + D_j)\}. \end{aligned}$$

We have now, applying (2.7) to the right member of (2.8),

$$E(Y_n)^k \leqq c_1 \sum_{j=0}^{\infty} (j+2)^k e^{-c_2 j} + \sum_{j=0}^{\infty} (j+2)^k \left(1 - H\left(\frac{jpd}{2}\right)\right). \tag{2.9}$$

The first series on the right of (2.9) obviously converges. Now consider the second. We have

$$\begin{aligned} \sum_{j=0}^{\infty} (j+2)^k \left(1 - H\left(\frac{jpd}{2}\right)\right) &= \sum_{j=0}^{\infty} (j+2)^k P\left\{R_1 > \frac{jpd}{2}\right\} \\ &\leqq \left(\frac{2}{pd}\right)^{k+1} E(R_1+2)^{k+1}. \end{aligned} \tag{2.10}$$

In [1] (relation (4.5)) it is shown that

$$\left((s-1)w_{ns} - \sum_{j=1}^{s-1} w_{nj}\right) \leqq Y_{n-1}$$

Hence (2.3) and the lemma follow for $k \geqq 1$. The proof for $0 < k < 1$ is almost the same; only a few obvious changes are needed in (2.8), (2.9), and (2.10).

3. Finiteness of m_k'. Of great interest is the question of when m_k is finite. In this section we shall give a sufficient condition for m_k' to be finite (and hence a fortiori for $m_k \leqq m_k'$ to be finite). We shall later see that this condition is essentially necessary for m_k' to be finite.

THEOREM 2. *If $k > 0$, and*

$$ER_1^{k+1} < \infty, \tag{3.1}$$

then

$$m_k' < \infty, \tag{3.2}$$

and

$$m_k < \infty. \tag{3.3}$$

PROOF. We assume that there exists a number $T > 0$ such that $g_1 < T$ with probability one. When we bear in mind how w_{n+1} is related to w_n, it follows immediately that, if Theorem 2 holds in this case, it a fortiori holds in general.

In order to carry out the proof we shall assume that $m_k' = \infty$ and obtain a contradiction. Let A be the set $\{x \mid x_1 < T\}$. Then from (2.2) we obtain that

$$\sup_n \int_A (x_s)^k \, dF_n(x) < \infty, \tag{3.4}$$

and hence

$$\sup_n \int_A (x_1 + \cdots x_s)^k \, dF_n(x) < \infty. \tag{3.5}$$

From the manner in which we obtain w_{n+1} from w_n we have that

$$W_{n+1} = W_n + R_n - sg_{n+1} \tag{3.6}$$

if $w_{n1} \geqq T$, and always we have

$$W_{n+1} \leqq W_n + R_n . \tag{3.7}$$

We now note the inequality (2.15.1) on page 39 of [7], which states that $r > 1$, $x \geqq 0$, $y \geqq 0$ imply that

$$x^r - y^r \leqq rx^{r-1}(x - y). \tag{3.8}$$

Putting $r = k + 1$, $x = W_{n+1}$, $y = W_n$, we have, from (3.6),

$$\begin{aligned} W_{n+1}^{k+1} - W_n^{k+1} &\leqq (k + 1)(W_n + R_n - sg_{n+1})^k(R_n - sg_{n+1}) \\ &= (k + 1)W_n^k\left\{\left(1 + \frac{R_n - sg_{n+1}}{W_n}\right)^k (R_n - sg_{n+1})\right\}. \end{aligned} \tag{3.9}$$

Consider the expression in brackets in the last expression of (3.9). By (3.1), the boundedness of g_{n+1}, and the independence of W_n from g_{n+1} and R_n, the conditional expected value of this bracketed expression, given W_n, tends to $E(R_n - sg_{n+1}) < 0$ as $W_n \to \infty$. Hence, if $EW_n^k \to \infty (= m_k')$ as $n \to \infty$, (3.9) implies that

$$\lim_{n\to\infty} E\{W_{n+1}^{k+1} - W_n^{k+1} \mid w_{n1} \geqq T\} = -\infty. \tag{3.10}$$

Similarly, putting $x = W_n + R_n$, $y = W_n$, and noting that $(a + b)^k \leqq 2^k(a^k + b^k)$ if $a, b, k \geqq 0$, (3.7) yields

$$\begin{aligned} W_{n+1}^{k+1} - W_n^{k+1} \leqq (W_n + R_n)^{k+1} - W_n^{k+1} &\leqq (k + 1)(W_n + R_n)^k R_n \\ &\leqq (k + 1)2^k(W_n^k + R_n^k)R_n . \end{aligned} \tag{3.11}$$

From (3.1), (3.5) and the independence of R_n and W_n, we conclude that there is a number $c < \infty$ such that

$$\sup_n E\{W_{n+1}^{k+1} - W_n^{k+1} \mid w_{n1} < T\} < c. \tag{3.12}$$

From (3.10), (3.12), and the fact that (3.5) and $m_k' = \infty$ imply that $\bar{A}$ has probability $>\epsilon > 0$ according to F_n for all sufficiently large n, we conclude that there is an integer N_0 such that $EW_{n+1}^{k+1} \leqq EW_n^{k+1}$ for $n \geqq N_0$. Since, for $n \leqq N_0$, $EW_n^{k+1} \leqq E(R_1 + \cdots R_{N_0})^{k+1} < \infty$, we conclude that $\sup_n EW_n^{k+1} < \infty$, contradicting the assumption that $m_k' = \infty$. This completes the proof.

4. Necessity of the condition (3.1). The present section is devoted to the proof of

THEOREM 3. *If, for any positive k,*

$$ER_1^{k+1} = \infty \tag{4.1}$$

and $Eg_1 < \infty$, then

$$m_k' = \infty . \tag{4.2}$$

It will easily be seen from our proof that Theorem 3 is a fortiori true if $\rho \geqq 1$. Only the case $\rho < 1$ requires proof and this is the case we shall consider.

PROOF. Let m be so large that

$$\int_M dF(x) = \alpha > 0$$

where M is the set of all points $(x_1, x_2, \cdots, x_s)$ in S such that $x_s \leqq m$. We have already remarked in Section 1 that the process $\{w_n^0\}$ there defined is stationary and metrically transitive. Let $\nu_1^0, \nu_2^0, \cdots$ be the indices n for which $w_n^0 \varepsilon M$, and define

$$\mu_i^0 = \nu_{i+1}^0 - \nu_i^0 .$$

It follows from the ergodic theorem that

$$E\mu_i^0 = \frac{1}{\alpha} < \infty .$$

Let $\{w_n'\}$ be the process obtained from $\{w_n\}$ as follows: $w_1' = w_1 = 0$. Thereafter $w_n' = w_n$ until the first index n, say ν_1', such that $w_{\nu_1'} \varepsilon M$; define $w_{\nu_1'}' = 0$. We now obtain each successive w_{n+1}' from its predecessor w_n' by using R_n and g_{n+1} in exactly the same manner as w_{n+1} is obtained from w_n, until the next index, say ν_2', for which $w_{\nu_2'}'$ would be in M; instead set $w_{\nu_2'}' = 0$. Continue in this manner to define $\{w_n'\}$. Define $\mu_i' = \nu_{i+1}' - \nu_i'$. Then $\mu_1', \mu_2', \cdots$ are independent, identically distributed chance variables. It follows from the construction of the process $\{w_n'\}$ and the first paragraph of Section 3 of [1] that $E\mu_i' \leqq E\mu_i^0$. Hence $E\mu_i'$ is finite. It follows from the strong law of large numbers that

$$P\left\{\lim_{n\to\infty} \frac{\nu_n'}{n} = E\mu_1'\right\} = 1. \tag{4.3}$$

We shall later show that

$$P\left\{\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} (w_{is}')^k = \infty\right\} = 1. \tag{4.4}$$

Since $w_n' \leqq w_n$ it follows at once that

$$P\left\{\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} (w_{is})^k = \infty\right\} = 1. \tag{4.5}$$

Hence

$$P\left\{\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} (W_i)^k = \infty\right\} = 1. \tag{4.6}$$

The desired result (4.2) follows from (1.16) and (4.6).

It remains to prove (4.4). Let $j(n)$ be defined for all integral n by

$$\nu_{j(n)}' \leqq n < \nu_{j(n)+1}' .$$

We shall later prove that

$$E\{(w'_{1s})^k + (w'_{2s})^k + \cdots + (w'_{\mu'_1,s})^k\} = \infty. \tag{4.7}$$

From this and the strong law of large numbers it follows that

$$P\left\{\lim_{n\to\infty} (j(n))^{-1} \sum_{i=1}^{\nu'_{j(n)}} (w'_{is})^k = \infty\right\} = 1. \tag{4.8}$$

From (4.3) and (4.8) we obtain that

$$P\left\{\lim_{n\to\infty} (\nu'_{j(n)})^{-1} \sum_{i=1}^{\nu'_{j(n)}} (w'_{is})^k = \infty\right\} = 1. \tag{4.9}$$

From (4.9) we have at once that

$$P\left\{\lim_{n\to\infty} (\nu'_{j(n)})^{-1} \sum_{i=1}^{n} (w'_{is})^k = \infty\right\} = 1. \tag{4.10}$$

Also

$$P\left\{\lim_{n\to\infty} \frac{n}{\nu'_{j(n)}} = 1\right\} = 1. \tag{4.11}$$

From (4.10) and (4.11) we have the desired result (4.4).

It remains to prove (4.7). Let N be an integer so large that

$$P\left\{\sum_{i=1}^{n} g_i < 2nEg_1 \text{ for all } n \geqq N\right\} > \tau > 0 \tag{4.12}$$

The existence of such an N follows from the strong law of large numbers. We may also assume N so large that $2NEg_1 > m$. Let $T = 4NEg_1$. Suppose that $t \geqq T$ and the largest integer contained in $(t/4Eg_1)$ is t'. Then $t' \geqq N$, and (4.12) implies that the conditional probability of the event A_1,

$$A_1 = \{\mu'_1 > t' \text{ and } w'_{ns} > 2t'Eg_1 \text{ for } 2 \leqq n \leqq t'\}, \tag{4.13}$$

given that $w'_{2s} = t$, is greater than τ. ($\mu'_1 > t'$ is implied by the other events in (4.13).) When the event A_1 occurs, we have

$$\sum_{n=1}^{\mu'_1} (w'_{ns})^k \geqq \sum_{n=1}^{t'} (w'_{ns})^k > t'(2t'Eg_1)^k \geqq ct^{k+1} \tag{4.14}$$

with $c > 0$. From (4.1) and the construction of the process $\{w'_n\}$ we have (by considering $((R_1 - g_2)^+)^{k+1}$ on the set where $g_2 < c$ where $c < \infty$ is chosen so that $P\{g_2 < c\} > 0$) that

$$E(w'_{2s})^{k+1} = \infty. \tag{4.15}$$

The desired result (4.7) follows from (4.14) and (4.15). This completes the proof of Theorem 3.

The following theorem can be proved in essentially the same manner as Theorem 3:

THEOREM 4. *If, for a positive integer N, an integer j ($1 \leqq j \leqq s$), and a positive k*

$$E(w_{Nj})^{k+1} = \infty, \tag{4.16}$$

then

$$\int (x_j)^k \, dF(x) = \infty. \tag{4.17}$$

Theorem 3 is a special case of Theorem 4 for the case $N = 2, j = s$. For then (4.1) implies (4.16), and (4.17) implies (4.2). Let M_i denote the ith smallest of $R_1, \cdots, R_s$, and suppose

$$E(M_i)^{k+1} = \infty. \tag{4.18}$$

Then (4.16) holds with $N = s, j = i$. This also implies Theorem 3, for (4.1) implies (4.18) for $i = s$. Finally we remark that (4.18) with $i = 1$ implies

$$m_k = \infty. \tag{4.19}$$

5. Implications for the one-dimensional random walk. The results of the preceding sections imply not only results on the behavior of queues in general, but also results on the random walk in s-dimensional space. We shall content ourselves with pointing out two of these implications for the one-dimensional random walk, although the results for the s-dimensional walk obtained in earlier sections are more general and usually more difficult to prove. Without further remark all problems treated in this section are to be assumed to be one-dimensional.

THEOREM 5. *Let $u_1, u_2, \cdots$ be independent, identically distributed chance variables. Let $S_n = \sum_{i=1}^{n} u_i$, and define*

$$v = \sup(0, S_1, S_2, S_3, \cdots).$$

If

$$-\infty \leqq Eu_1 < 0, \tag{5.1}$$

and, for $k > 0$,

$$E(u_1^+)^{k+1} < \infty, \tag{5.2}$$

then

$$Ev^k < \infty. \tag{5.3}$$

THEOREM 6. *With the definitions of Theorem 5, if*

$$-\infty < Eu_1 < 0, \tag{5.4}$$

and, for $k > 0$,

$$E(u_1^+)^{k+1} = \infty, \tag{5.5}$$

then

$$Ev^k = \infty. \tag{5.6}$$

PROOF. Consider the process: $w_1^* = u_1^+$, $w_{n+1}^* = (w_n^* + u_{n+1})^+$, $n \geqq 1$. Let $F_n^*(z)$ be the d.f. of w_n^*, and let

$$F^*(z) = \lim_{n\to\infty} F_n^*(z)$$

when the latter exists. It was shown in [3] and follows from the results of [1] for the case $s = 1$ that, when $u_n = R_n - g_{n+1}$, $F^*(z)$ exists, is a distribution function, and equals the limiting d.f. $F(z)$ of w_n. It was also shown in [3] that the distribution function of v is then $F^*(z)$. An examination of the proofs of these statements shows that they are valid for the process $\{w_n^*\}$ even when u_n is not of the form $R_n - g_{n+1}$, provided only that (5.1) is satisfied. An examination of the proofs of Section 1 and Theorem 3 of the present paper shows that they too hold even if u_n is not of the form $R_n - g_{n+1}$. But then Theorem 6 is simply a restatement of Theorem 3.

It is sufficient to prove Theorem 5 for chance variables $\{u_n^*\}$, where $u_n^* = \max(u_n, -T)$ and $T > 0$ is so large that $Eu_n^* < 0$. But $u_n^* = (u_n^* + T) - T$ and is therefore of the form $R_n - g_{n+1}$, with $R_n = (u_n^* + T)$, $g_{n+1} \equiv T$. Theorem 5 is then simply a restatement of Theorem 2.

While the results of the present paper on the queueing process and the corresponding s-dimensional random walk are new, Theorems 5 and 6 on the one-dimensional random walk were also obtained by Darling, Erdös, and Kakutani, to whom the problem was communicated by us. These writers also obtained other related results, and they have informed us that many of these results are implicit in [4]. In the course of the present work we have had interesting discussions with Professor Shizuo Kakutani.

6. The mean queue length. As in [1], Section 9, let Q_i be the number of individuals in the queue waiting to be served, just before the service of the ith individual begins. To avoid trivial circumlocutions we assume $G(0) = 0$ ($G(x)$ is the d.f. of g_i). In [1] the limit $D(x)$ of $D_n(x)$, the d.f. of Q_n, is shown to exist and $D(x)$ is explicitly given. We shall now be concerned with

$$\bar{Q}_n = n^{-1} \sum_{i=1}^{n} Q_i .$$

Let $\{w_n^0\}$ be the process defined in Section 1. We now construct a process $\{w_n^0, Q_n^0\}$, where $Q_1^0, Q_2^0, \cdots$ remain to be defined. Let $t_n = \sum_{i=1}^{n} g_i$. We define

Q_n^0 to be equal to the number of indices i which satisfy

$$t_n < t_i \leqq t_n + w_{n1}^0 . \tag{6.1}$$

It follows that the process $\{w_n^0, Q_n^0\}$ is stationary and metrically transitive, so that, by the ergodic theorem, $\bar{Q}_n^0 = n^{-1} \sum_{i=1}^n Q_i^0$ approaches a constant limit c,

$$c = \int x \, dD(x),$$

with probability one. (It is easy to prove that c is contained between $Ew_{n1}^0/Eg_1 - 1$ and Ew_{n1}^0/Eg_1.) Since $w_{n1} \leqq w_{n1}^0$ it follows from (6.1) that $Q_n \leqq Q_n^0$. Hence

$$P\{\limsup_n \bar{Q}_n \leqq c\} = 1. \tag{6.2}$$

Just as in Section 1, one proves that

$$P\{\liminf_n \bar{Q}_n \geqq c\} = 1. \tag{6.3}$$

Hence

$$P\{\lim_n \bar{Q}_n = c\} = 1. \tag{6.4}$$

7. The duration of busy periods. A busy period is a closed time interval, say $t' \leqq t \leqq t''$, such that all s servers are occupied throughout this interval, $t'' - t' > 0$, and the interval is maximal, i.e., if $\tau' \leqq t' < t'' \leqq \tau''$, $\tau'' - \tau' > t'' - t'$, then all s servers are not occupied for some time point in the interval (τ', τ''). The length of the busy period is $t'' - t'$, t' is its beginning, and t'' is its end. Let B_i be the sum of the lengths of all busy periods at or before t_i; if t_i is in the interior of a busy period, we count into B_i the length of the interval from the beginning of the period until t_i.

It is easy to verify that whether or not any time point t with $t_i < t < t_{i+1}$ is in a busy interval depends only on w_i, R_i, and g_{i+1}. Since the value of B_n is unaffected by removing from busy periods any of the points t_i $(1 \leqq i \leqq n)$ contained in them, it follows that the process

$$\{B_n, w_n\}, n = 1, 2, \cdots$$

is Markoffian.

Let $\{w_n^0\}$ be the process defined in Section 1. Define $B_1^0 = 0$. Define B_n^0, $n \geqq 2$, to be the same function of the process $\{w_n^0\}$ as B_n is of the process $\{w_n\}$. Since $w_n^0 \geqq w_n$ with probability one, it follows that $B_n^0 \geqq B_n$ with probability one.

Since the process $\{w_n^0\}$ is stationary and metrically transitive, so is the process

$$\{B_{n+1}^0 - B_n^0\}, n = 1, 2, \cdots .$$

Hence

$$P\left\{\lim \frac{B_n^0}{n} = E(B_2^0)\right\} = 1.$$

In essentially the same manner as in Section 1 one proves easily that

$$P\left\{\lim \frac{B_n}{n} = E(B_2^0)\right\} = 1.$$

From this we obtain immediately that

$$P\left\{\lim \frac{B_n}{t_n} = \frac{E(B_2^0)}{Eg_1}\right\} = 1.$$

This gives the long-term average time spent in busy periods.

The limiting distribution of the length of a busy period can be obtained in a very tedious but straightforward manner from the marginal distributions of the process $\{w_n^0\}$.

REFERENCES

[1] J. Kiefer, and J. Wolfowitz, "On the theory of queues with many servers," *Trans. Amer. Math. Soc.*, Vol. 78, 1, January 1955, pp. 1–18.

[2] J. V. Uspensky, "Introduction to mathematical probability," McGraw-Hill Book Company, Inc., New York, 1937.

[3] D. V. Lindley, "The theory of queues with a single server," *Proc. Cambridge Philos. Soc.*, Vol. 48 (1952), Part 2, pp. 277–89.

[4] P. Erdös, "On a theorem of Hsu and Robbins," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 286–291.

[5] J. Wolfowitz, "The efficiency of sequential estimates etc.," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.

[6] J. L. Doob, "Stochastic processes," John Wiley & Sons, Inc., New York, 1953.

[7] G. H. Hardy, J. E. Littlewood, and G. Polya, "Inequalities," Cambridge University Press, Cambridge, 1934.

TOLERANCE REGIONS

BY D. A. S. FRASER AND IRWIN GUTTMAN

University of Toronto

1. Summary. In this paper definitions are given for three types of tolerance regions. For distribution-free tolerance regions, an analytic condition is derived for the characteristic function of the region. Examples of the application of the condition are considered. For β-expectation tolerance regions, a criterion for a good tolerance region is introduced, and it is shown that the problem of finding such a tolerance region can be reduced to that of finding a good test for an equivalent hypothesis-testing problem. Best tolerance regions are obtained for a number of single variate and multivariate problems involving normal distributions.

2. Introduction. Let $\mathfrak{X}$ $(\mathfrak{A})$ be a measurable space and $\{P_X^\theta \mid \theta \varepsilon \Omega\}$ be a class of probability measures defined over $\mathfrak{X}(\mathfrak{A})$. For the theory in this paper we assume that an experiment corresponds to a sample of n from a component experiment. Hence our sample space is $\mathcal{W} = \mathfrak{X}^n$, and the probability measures are the nth power product of the measures $\{P_X^\theta \mid \theta \varepsilon \Omega\}$. We designate these measures by $\{P_W^\theta \mid \theta \varepsilon \Omega\}$.

A statistical tolerance region is a mapping from the sample space $\mathcal{W}$ to the space of subsets $\mathfrak{A}$ of the component space.

DEFINITION 2.1. A statistical tolerance region, $S(x_1, \cdots, x_n)$, is a statistic defined over $\mathcal{W} = \mathfrak{X}^n$ and taking values in the σ-algebra $\mathfrak{A}$.

In application the statistician calculates from the outcome $(x_1, \cdots, x_n)$ a region $S(x_1, \cdots, x_n)$ in the space $\mathfrak{X}$ which is being sampled, and then makes some probability or expectation statement about the probability measure of this set.

We first consider distribution-free tolerance regions. Heretofore the term "nonparametric" has generally been applied to these regions, but in accordance with the use of the term "distribution free" in other branches of statistics, and because these regions can also be considered for parametric problems, we prefer the term distribution free.

DEFINITION 2.2. $S(x_1, \cdots, x_n)$ is a distribution-free tolerance region for $\{P_X^\theta \mid \theta \varepsilon \Omega\}$ if the induced probability distribution of

$$P_X^\theta(S(x_1, \cdots, x_n))$$

corresponding to the measure P_W^θ over $\mathfrak{X}^n$ is independent of the parameter $\theta \varepsilon \Omega$.

Because the probability measure or coverage of a distribution-free tolerance region has a "known" distribution independent of the "unknown" parameter,

Received December 16, 1954.

the statistician is able to make a probability statement about the coverage of the region.

The next definition is proposed more with a view toward the immediate requirements of a statistician.

DEFINITION 2.3. $S(x_1, \cdots, x_n)$ is a β-content tolerance region at confidence level C if

$$\Pr_\theta\{P_X^\theta(S(X_1, \cdots, X_n)) \geqq \beta\} \geqq C$$

for all $\theta \varepsilon \Omega$.

For such a region, the statistician has confidence C that the probability content of the region $S(x_1, \cdots, x_n)$ is at least β, regardless of the measure being sampled. Of course in some situations he may prefer that $S(x_1, \cdots, x_n)$ satisfy the relation

$$\Pr_\theta\{\beta_1 \leqq P_X^\theta(S(X_1, \cdots, X_n)) \leqq \beta_2\} \geqq C$$

for all $\theta \varepsilon \Omega$.

The next type of tolerance region has had perhaps less attention from the applied statistician than it deserves.

DEFINITION 2.4. $S(x_1, \cdots, x_n)$ is a β-expectation tolerance region if

$$E_W^\theta\{P_X^\theta(S(X_1, \cdots, X_n))\} \leqq \beta \qquad \text{for all } \theta \varepsilon \Omega.$$

For such a region the average probability content of the region is at most β.

In hypothesis testing the reduction to similarity is sometimes helpful for finding a whole class of tests in convenient form. For tolerance regions, we therefore propose the following definition:

DEFINITION 2.5. $S(x_1, \cdots, x_n)$ is a similar β-expectation tolerance region if

$$E_W^\theta\{P_X^\theta(S(X_1, \cdots, X_n))\} = \beta$$

for all $\theta \varepsilon \Omega$.

A similar β-expectation region can also be viewed as a β-confidence region for a future observation from the distribution being sampled. For by noting that $P_X^\theta(S(x_1, \cdots, x_n))$ is the probability that another observation falls in S given $x_1, \cdots, x_n$, we see that the left-hand side of the expression in Definition 2.5 is the marginal probability of such an event. This probability is equal to β; hence there is β confidence that the future observation falls in S.

3. Distribution-free tolerance regions. For distribution-free tolerance regions we are able to give a necessary and sufficient analytic condition. To do this we need the definition of a characteristic function, $\varphi_y(x_1, \cdots, x_n)$, of a region $S(x_1, \cdots, x_n)$:

$$\begin{aligned} \varphi_y(x_1, \cdots, x_n) &= 1 \qquad \text{if } y \varepsilon S(x_1, \cdots, x_n) \\ &= 0 \qquad \text{if } y \not\varepsilon S(x_1, \cdots, x_n) \end{aligned} \tag{3.1}$$

where $y \varepsilon \mathfrak{X}$. Then it is easily seen that

$$P_X^\theta(S(x_1, \cdots, x_n)) = E_Y^\theta\{\varphi_Y(x_1, \cdots, x_n)\}$$

where the expectation applies to the random variable Y with probability measure P_X^θ.

THEOREM 3.1. *A necessary and sufficient condition that $S(x_1, \cdots, x_n)$ be a distribution-free tolerance region is that there exist a sequence of real numbers $\alpha_1, \alpha_2, \cdots$ such that*

$$\varphi_{y_1}(x_1, \cdots, x_n) - \alpha_1, \qquad \varphi_{y_1}(x_1, \cdots, x_n)\varphi_{y_2}(x_1, \cdots, x_n) - \alpha_2, \cdots$$

are respectively unbiased estimates of zero over $\mathfrak{X}^{n+1}, \mathfrak{X}^{n+2}, \cdots$ for the power product measures of $\{P_X^\theta \mid \theta \varepsilon \Omega\}$. The sequence $\alpha_1, \alpha_2, \cdots$ is the moment sequence for the distribution of $V = P_X^\theta(S(X_1, \cdots, X_n))$, where the X_i have measure P_X^θ.

PROOF. A distribution-free tolerance region has the distribution function, say $F_\theta(v)$, independent of θ. Now, since a distribution function on a bounded interval is uniquely determined by the corresponding moment sequence and conversely (see [1]), it is equivalent to state that the moment sequence for $F_\theta(v)$ is independent of θ.

Letting α_r be the rth moment of $F_\theta(v)$, then

$$\begin{aligned}
\alpha_r &= \int_0^1 v^r\, dF_\theta(v) \\
&= \int_{\mathfrak{X}^n} [P_X^\theta(S(x_1, \cdots, x_n))]^r \prod_{i=1}^n dP_X^\theta(x_i) \\
&= \int_{\mathfrak{X}^n} [E_Y^\theta\{\varphi_Y(x_1, \cdots, x_n)\}]^r \prod_{i=1}^n dP_X^\theta(x_i) \\
&= \int_{\mathfrak{X}^n} \left[\int_{\mathfrak{X}} \varphi_y(x_1, \cdots, x_n)\, dP_X^\theta(y)\right]^r \prod_{i=1}^n dP_X^\theta(x_i) \\
&= \int_{\mathfrak{X}^{n+r}} \prod_{i=1}^{i} \varphi_{y_i}(x_1, \cdots, x_n) \prod_{j=1}^r dP_X^\theta(y_j) \prod_{i=1}^n dP_X^\theta(x_i).
\end{aligned}$$

Therefore, $\prod_{j=1}^r \varphi_{y_j}(x_1, \cdots, x_n) - \alpha_r$ is an unbiased estimate of zero over $\mathfrak{X}^{n+r}$. Thus, the statement that $F_\theta(v)$ is independent of θ is equivalent to the existence of the sequence $\alpha_1, \alpha_2, \cdots$ such that the above expression is an unbiased estimate of zero for all r.

For some theoretical developments it is convenient to have a definition of a randomized tolerance region. Let Z be a random variable whose probability measure is a measurable function of $x_1, \cdots, x_n$.

DEFINITION 3.1. $S(x_1, \cdots, x_n; z)$ is a randomized distribution-free tolerance region for $\{P_X^\theta \mid \theta \varepsilon \Omega\}$ if the induced probability distribution of $P_X^\theta(S(x_1, \cdots, x_n; z))$, corresponding to the measure P_W^θ over $\mathfrak{X}^n$ and the random variable Z for z, is independent of the parameter θ. It is assumed that $S(x_1, \cdots, x_n; z)$ is a measurable function of $(x_1, \cdots, x_n, z)$.

As for the nonrandomized case, we define a function

$$\Phi_y(x_1, \cdots, x_n; z) = 1 \quad \text{if } y \in S(x_1, \cdots, x_n; z) \tag{3.2}$$
$$= 0 \quad \text{if } y \notin S(x_1, \cdots, x_n; z).$$

Taking the expectation with respect to Z, we define a related function

$$\varphi_{y_1 \cdots y_r}(x_1, \cdots, x_n) = E_Z\left\{\prod_{j=1}^{r} \Phi_{y_j}(x_1, \cdots, x_n; Z)\right\} \tag{3.3}$$

a function which is characteristic of the tolerance region. Then we have the extension of Theorem 3.1:

THEOREM 3.2. *A necessary and sufficient condition that $S(x_1, \cdots, x_n; z)$ be a distribution-free randomized tolerance region is that there exist a sequence of real numbers $\alpha_1, \alpha_2, \cdots$ such that $\varphi_{y_1}(x_1, \cdots x_n) - \alpha_1$, $\varphi_{y_1 y_2}(x_1, \cdots, x_n) - \alpha_2, \cdots$ are respectively unbiased estimates of zero over $\mathfrak{X}^{n+1}, \mathfrak{X}^{n+2}, \cdots$ for the power product measures of $\{P_X^\theta \mid \theta \in \Omega\}$. The sequence $\alpha_1, \alpha_2, \cdots$ is the moment sequence for the distribution of $V = P_X^\theta(S(X_1, \cdots, X_n; Z))$ where the X_i have the measure P_X^θ.*

PROOF. This follows the method of proof given for Theorem 3.1.

We give now some examples of the application of the above theorem for the nonrandomized case.

EXAMPLE 3.1. Consider sampling from an arbitrary discrete distribution on the real line. We have $W = R^n$, and the class of probability measures is $\{P_X^\theta \mid \theta \in \Omega\}$, where θ here indexes the discrete distributions on R^1. We establish that there do not exist distribution-free tolerance regions $S(x_1, \cdots, x_n)$ symmetric in the x's, other than the trivial tolerance region $S = \emptyset$ or $\mathfrak{X}$.

Let $S(x_1, \cdots, x_n)$ be a distribution-free tolerance region which is symmetric in the x's. We show that either $S(x_1, \cdots, x_n) = \emptyset$ or $S(x_1, \cdots, x_n) = \mathfrak{X}^n$.

If $\varphi_y(x_1, \cdots, x_n)$ is the characteristic function of $S(x_1, \cdots, x_n)$, then by Theorem 3.1 we have the existence of $\alpha_1, \alpha_2, \cdots$ such that

$$\prod_{j=1}^{r} \varphi_{y_j}(x_1, \cdots, x_n) - \alpha_r \tag{3.4}$$

is an unbiased estimate of zero over $\mathfrak{X}^{n+r}$.

For samples from $\mathfrak{X}^n$ we define a statistic called the order statistic

$$t(x_1, \cdots, x_n) = \{x_1, \cdots, x_n\}.$$

This statistic gives the values of the x's in the outcome $(x_1, \cdots, x_n)$, but not the order in which they occur in this outcome. Now it is easily shown that for the class of power product measures over $\mathfrak{X}^n$ considered here, this statistic is sufficient. Halmos [2] has shown that $t(x_1, \cdots, x_n)$ is complete for the measures above.

We have that (3.4) is an unbiased estimate of zero:

$$E_\theta\left\{\prod_{j=1}^{r} \varphi_{X_{n+j}}(X_1, \cdots, X_n) - \alpha_r\right\} = 0. \tag{3.5}$$

Since $t(x_1, \cdots, x_{n+r})$ is a sufficient statistic, the expression

$$E\left\{\prod_{j=1}^{r} \varphi_{X_{n+j}}(X_1, \cdots, X_n) - \alpha_r \,\middle|\, t(\underline{X}) = t\right\}$$

is independent of θ, that is, it is a statistic. But (3.5) can be written as

$$E_\theta\left\{E\left\{\prod_{j=1}^{i} \varphi_{X_{n+j}}(X_1, \cdots, X_n) - \alpha_r \,\middle|\, t(\underline{X}) = T\right\}\right\} = 0,$$

where the first expectation operator applies to the induced distribution of $t(x_1, \cdots, x_{n+r})$. From completeness over $\mathfrak{X}^{n+r}$, we have

$$E\left\{\prod_{j=1}^{r} \varphi_{X_{n+j}}(X_1, \cdots, X_n) - \alpha_r \,\middle|\, t(\underline{X}) = t\right\} = 0 \tag{3.6}$$

almost everywhere with respect to the induced measures of $t(x_1, \cdots, x_{n+r})$. Since the class $\{P_X^\theta \mid \theta \varepsilon \Omega\}$ is the class of all discrete distributions, almost everywhere means everywhere.

We consider (3.6) with $r = 1$. The conditional distribution given the statistic $t(x_1, \cdots, x_{n+1})$ gives equal probability to all permutations of $(x_1, \cdots, x_{n+1})$. Hence (3.6) with $r = 1$ becomes

$$\frac{1}{(n+1)!} \sum_P \varphi_{x_{i_{n+1}}}(x_{i_1}, \cdots, x_{i_n}) - \alpha_1 = 0 \tag{3.7}$$

everywhere; P designates summation with respect to all permutations $i_1, \cdots, i_{n+1}$ of $(1, \cdots, n+1)$. Since $S(x_1, \cdots, x_n)$ is symmetric in the x's, so also is $\varphi_y(x_1, \cdots, x_n)$ symmetric in $x_1, \cdots, x_n$. Therefore (3.7) becomes

$$\begin{aligned} \varphi_{x_{n+1}}(x_1, \cdots, x_n) + \varphi_{x_n}(x_1, \cdots, x_{n-1}, x_{n+1}) + \cdots + \varphi_{x_1}(x_2, \cdots, x_{n+1}) \\ = (n+1)\alpha_1, \end{aligned} \tag{3.8}$$

and (3.8) holds for all $x_1, \cdots, x_{n+1}$. Taking $x_1 = x_2 = \cdots = x_{n+1} = x$, we have

$$(n+1)\varphi_x(x, \cdots, x) = (n+1)\alpha_1 .$$

The quantity $\varphi_x(x, \cdots, x)$ can be either zero or one. Hence $(n+1)\alpha_1 = 0$ or $(n+1)$; that is, $\alpha_1 = 0$ or $= 1$. Thus the first moment of a random variable restricted to the interval 0, 1 is either zero or one. Obviously, the random variable (the coverage of the tolerance region) takes the value zero or one with probability one. Because the class of measures is the class of discrete distributions, this means either that

$$S(x_1, \cdots, x_n) = \emptyset$$

or that

$$S(x_1, \cdots, x_n) = \mathfrak{X}.$$

EXAMPLE (3.2) Consider sampling from an arbitrary absolutely continuous distribution on the real line. We have $\mathfrak{X} = R$ and we let θ in $\{P_X^\theta \mid \theta \in \Omega\}$ index the absolutely continuous distributions. We find the form of distribution-free upper tolerance limits in special cases. Suppose a distribution-free tolerance region $S(x_1, \cdots, x_n)$ has the form

$$(3.9) \qquad S(x_1, \cdots, x_n) =]-\infty, u(x_1, \cdots, x_n)].$$

(Intervals are open at the end where a reversed bracket appears, and closed otherwise.) Then $u(x_1, \cdots, x_n)$ is called a distribution-free upper tolerance limit. We assume that $u(x_1, \cdots, x_n)$ is symmetric in $x_1, \cdots, x_n$.

For convenience we define L_r to be the Lebesgue measure over R^r. As in Example 3.1, it can easily be shown that the order statistic is sufficient for the class of absolutely continuous distributions. It has been proved complete by Lehmann [3] and Fraser [4]. Following the argument in Example 3.1, we obtain from Theorem 3.1 with $r = 1$ that

$$\frac{1}{(n+1)} \sum_P \varphi_{x_{i_{n+1}}}(x_{i_1}, \cdots, x_{i_n}) = \alpha_1$$

almost everywhere (Lebesgue) over R^{n+1}. Since $\varphi_y(x_1, \cdots, x_n)$ is symmetric in the x's, we have

$$(3.10) \qquad \varphi_{x_{n+1}}(x_1, \cdots, x_n) + \cdots + \varphi_{x_1}(x_2, \cdots, x_{n+1}) = (n+1)\alpha_1$$

almost everywhere. Because φ is a characteristic function, $(n+1)\alpha_1$ is one of the integers $0, 1, \cdots, n, n+1$. We find the form of $u(x_1, \cdots, x_n)$ when $(n+1)\alpha_1$ is 0, 1, n, or $n+1$.

Consider the case $(n+1)\alpha_1 = 1$. We shall prove that

$$\begin{aligned} \varphi_y(x_1, \cdots, x_n) &= 1 \qquad \text{if } y \leqq x_{(1)} \\ &= 0 \qquad \text{if } y > x_{(1)} \end{aligned}$$

almost everywhere in R^{n+1}. In terms of $u(x_1, \cdots, x_n)$, this means that

$$u(x_1, \cdots, x_n) = x_{(1)}$$

almost everywhere in R^n. Let $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ designate the numbers $x_1, \cdots, x_n$ arranged in order of increasing magnitude: $x_{(1)} \leqq \cdots \leqq x_{(n)}$.

Suppose $\varphi_y\ (x_1, \cdots, x_n) = 1$ on a set of positive Lebesgue measure in the region of R^{n+1} for which $y > x_{(1)}$. From the properties of measure, it follows that there exists a positive δ such that $\varphi_y(x_1, \cdots, x_n) = 1$ on a set of positive measure in the region $y > x_{(1)} + \delta$; call this set A. Divide the space R^n of $(x_1, \cdots, x_n)$ into "cubes" having sides of length ϵ, a typical set being

$$(3.11) \qquad \{(x_1, \cdots, x_n) \mid m_i\, \varepsilon < x_i \leqq (m_i + 1)\, \varepsilon; \quad (i = 1, \cdots, n)\}.$$

We let, of course, m_i for each i range over all real integers. There is a countable number of such sets. Consider the following set B:

$$B = \{(x_1, \cdots, x_n) \mid L_1\{y \mid (x_1, \cdots, x_n, y) \,\varepsilon\, A\} > 0\}.$$

From the properties of measure, there exists at least one of the above-defined cubes which intersects B on a set of positive measure. For a later purpose we require $\epsilon < \delta$.

Now by choosing ϵ sufficiently small we can ensure the existence of at least one cube which intersects B on a set of positive measure and at the same time is disjoint from each of the diagonal sets,

$$\{(x_1, \cdots, x_n) \mid x_i = x_j\}, \tag{3.12}$$

which have measure zero. Designate one of these cubes by C. We summarize the results so far. For $(x_1, \cdots, x_n) \varepsilon B \cap C$, we have $\varphi_y(x_1, \cdots, x_n) = 1$ at least for y belonging to a set of positive measure in $y > x_{(1)} + \delta$.

From the definition of $\varphi_y(x_1, \cdots, x_n)$ we know it is monotone nonincreasing in y and takes the values 0, 1. That is, if $\varphi_{y^*}(x_1, \cdots, x_n) = 1$, then $\varphi_y(x_1, \cdots, x_n) = 1$ for $y < y^*$. Now from the last statement in the above paragraph, it follows that if $(x_1, \cdots, x_n) \varepsilon B \cap C$, then $\varphi_y(x_1, \cdots, x_n) = 1$ for $y \leqq x_{(1)} + \delta$.

We now derive a contradiction to (3.10). Without loss of generality, let x_1 be the smallest of the co-ordinates for points in C (it will always be the same co-ordinate because C does not intersect the diagonal sets (3.12.)). Consider the set D in R^{n+1}:

$$D = \{(x_1, \cdots, x_{n+1}) \mid (x_1, \cdots, x_n) \varepsilon C \cap B \quad ; \quad (x_{n+1}, x_2, \cdots, x_n) \epsilon C \cap B\}.$$

From the first condition defining D, we have $\varphi_{x_{n+1}}(x_1, \cdots, x_n) = 1$ for $x_{n+1} \leqq x_1 + \delta$. From the second condition defining D, we have $\varphi_{x_1}(x_{n+1}, x_2, \cdots, x_n) = 1$ for $x_1 \leqq x_{n+1} + \delta$. But for $(x_1, \cdots, x_{n+1}) \varepsilon D$ we have x_1 and x_{n+1} both as possible first co-ordinates for a point in C; hence $|x_1 - x_{n+1}| < \varepsilon$. Therefore if $(x_1, \cdots, x_{n+1}) \varepsilon D$, $\varphi_{x_{n+1}}(x_1, \cdots, x_n) = 1 = \varphi_{x_1}(x_{n+1}, x_2, \cdots, x_n)$, since the two conditions above are fulfilled by reason of our choice of $\epsilon < \delta$. For $(x_1, \cdots, x_{n+1}) \varepsilon D$ we have the left-hand side of (3.10) equal to at least 2, while the right-hand side of (3.10) by assumption was 1. This is a contradiction if we show D has positive measure.

Let L_n designate Lebesgue measure in R^n and let $\psi(x_1, \cdots, x_n)$ be the characteristic function of $B \cap C$ in R^n.

$$\begin{aligned} L_{n+1}(D) &= \int_{R^{n+1}} \psi(x_1, \cdots, x_n)\psi(x_{n+1}, x_2, \cdots, x_n) \prod_{i=1}^{n+1} dL_1(x_i) \\ &= \int_{R^{n-1}} \int_R \psi(x_1, \cdots, x_n)\, dL_1(x_1) \int_R \psi(x_{z+1}, \cdots, x_n)\, dL_1 \\ &\quad \times (x_{n+1}) \prod_{j=2}^{n} dL_1(x_j) \\ &= \int_{R^{n-1}} M^2(x_2, \cdots, x_n) \prod_{j=2}^{n} dL_1(x_j). \end{aligned} \tag{3.13}$$

But

$$L_n(B \cap C) = \int_{R^n} \psi(x_1, \cdots, x_n) \prod_{i=1}^{n} dL_1(x_i) \tag{3.14}$$
$$= \int_{R^{n-1}} M(x_2, \cdots, x_n) \prod_{i=2}^{n} dL_1(x_i).$$

Since L_n $(B \cap C) > 0$ by construction of $B \cap C$, then by comparing (3.13) and (3.14) we obtain $L_{n+1}(D) > 0$. This is the contradiction we worked toward. Therefore our assumption that $\varphi_y(x_1, \cdots, x_n) = 1$ on a set of positive measure in the region of R^{n+1} for which $y > x_{(1)}$ was false.

We have that $\varphi_y(x_1, \cdots, x_n) = 1$ on a set of positive measure only if that set is in $y \leqq x_{(1)}$. Thus for $x_{i_1} < \cdots < x_{i_{n+1}}$ there is only one term of (3.10) which can be 1 on a set of positive measure. However the right-hand side of (3.10) by assumption was 1 almost everywhere over $x_{i_1} < \cdots < x_{i_{n+1}}$. Therefore $\varphi_{x_{i_1}}(x_{i_2}, \cdots, x_{i_{n+1}}) = 1$ almost everywhere when $x_{i_1} < \cdots < x_{i_{n+1}}$. That is,

$$\begin{aligned} \varphi_y(x_1, \cdots, x_n) &= 1 \qquad \text{if } y \leqq x_{(1)} \\ &= 0 \qquad \text{if } y > x_{(1)}, \end{aligned}$$

and the distribution-free upper tolerance bound $u(x_1, \cdots, x_n)$ equals $x_{(1)}$.

Similarly if $(n + 1)\alpha_1 = n$, then $u(x_1, \cdots, x_n) = x_{(n)}$, and by an almost trivial argument $u(x_1, \cdots, x_n) = -\infty, +\infty$ according as $(n + 1)\alpha_1 = 0$, $n + 1$.

4. β-Content tolerance regions. Any tolerance region satisfying Definition 2.1 will produce a β-content tolerance region for suitably chosen C; for example,

$$C = \inf_{\theta \epsilon \Omega} \Pr_\theta\{P_X^\theta(S(X_1, \cdots, X_n)) \geqq \beta\}.$$

Also, a distribution-free tolerance region will produce a β-content tolerance region with a property of similarity. For if $S(x_1, \cdots, x_n)$ satisfies Definition 2.2, then, letting C equal the expression

$$\Pr_\theta\{P_X^\theta(S(X_1, \cdots, X_n)) \geqq \beta\},$$

which does not depend on θ, we have a similar β-content tolerance region given by

$$\Pr_\theta\{P_X^\theta(S(X_1, \cdots, X_n)) \geqq \beta\} = C.$$

5. β-expectation tolerance regions. First we prove some general properties of β-expectation tolerance regions. In Section 3 we defined by Formula 3.1 the characteristic function $\varphi_y(x_1, \cdots, x_n)$ of a nonrandomized tolerance region, and by Formula 3.3 with $r = 1$ we defined a characteristic function $\varphi_y(x_1, \cdots, x_n)$ of a randomized tolerance region. As a converse, we have the

THEOREM 5.1. *If $\varphi_y(x_1, \cdots, x_n)$ is a measurable function with* $0 \leqq \varphi_y(x_1, \cdots, x_n) \leqq 1$, *then there exists a tolerance region $S(x_1, \cdots, x_n)$ having $\varphi_y(x_1, \cdots, x_n)$ as its characteristic function.*

PROOF. Let Z be a random variable which has the uniform distribution on [0, 1] and define a randomized tolerance region by

$$S'(x_1, \cdots, x_n; z) = \{y \mid \varphi_y(x_1, \cdots, x_n) \geqq z\}.$$

Now we calculate the characteristic function of $S'(x_1, \cdots, x_n; z)$ and using (3.2) obtain

$$\begin{aligned}\varphi'(x_1, \cdots, x_n) &= E_Z\{\Phi_y'(x_1, \cdots, x_n; Z)\}\\ &= \Pr_Z\{\varphi_y(x_1, \cdots, x_n) \geqq Z\}\\ &= \varphi_y(x_1, \cdots, x_n).\end{aligned}$$

This proves the theorem.

We also state a theorem on similar β-expectation tolerance regions.

THEOREM 5.2. *A necessary and sufficient condition that $S(x_1, \cdots, x_n; z)$ be a similar β-expectation tolerance region is that $\varphi_y(x_1, \cdots, x_n) - \beta$ be an unbiased estimate of zero for the power product measure of P_X^θ over $\mathfrak{X}^{n+1}$.*

PROOF. Let $S(x_1, \cdots, x_n)$ be a tolerance region; then the expected content is

$$E_{WZ}^\theta\{P_X^\theta(S(X_1, \cdots, X_n; Z))\}. \tag{5.1}$$

From the definition of $\varphi_y(x_1, \cdots, x_n)$ this becomes

$$E_{WY}^\theta\{\varphi_y(X_1, \cdots, X_n)\}. \tag{5.2}$$

Obviously, then, a necessary and sufficient condition that (5.1) be equal to β is that $\varphi_y(x_1, \cdots, x_n) - \beta$ be an unbiased estimate of zero.

To introduce the notion of a good tolerance region, we need a function which gives us for each θ in Ω the relative merits of sets S in $\mathfrak{A}$. Let the "desirability" of a set S when the probability measure is P_X^θ be given by a probability measure $Q_\theta(S)$ defined for all $S \in \mathfrak{A}$. Then for a tolerance region S we define the power to be

$$E_W^\theta\{Q_\theta(S(X_1, \cdots, X_n))\}; \tag{5.3}$$

it is the average value of the "desirability" of the set S and in general is a function of θ. In terms of the characteristic function of $S(x_1, \cdots, x_n)$, the β-expectation condition is

$$\int_{\mathfrak{X}^{n+1}} \varphi_y(x_1, \cdots, x_n)\, dP_X^\theta(y) \prod_{i=1}^n dP_X^\theta(x_i) \leqq \beta, \tag{5.4}$$

and the power is

$$\int_{\mathfrak{X}^{n+1}} \varphi_y(x_1, \cdots, x_n)\, dQ_\theta(y) \prod_{i=1}^n dP_X^\theta(x_i). \tag{5.5}$$

The problem of finding a good tolerance region is then to find a characteristic function satisfying the size condition (5.4) and having good properties for the power (5.5). Obviously, this is equivalent to finding a good test function $\varphi_y(x_1, \cdots, x_n)$ for the hypothesis testing problem, over $\mathfrak{X}^{n+1}$,

$$\text{(5.6)}\qquad \begin{array}{lll} \text{Hypothesis:} & (P_X^\theta, \cdots, P_X^\theta, P_X^\theta), & \theta \,\varepsilon\, \Omega; \\ \text{Alternative:} & (P_X^\theta, \cdots, P_X^\theta, Q_\theta), & \theta \,\varepsilon\, \Omega; \end{array}$$

$(P_X^\theta, \cdots, P_X^\theta, Q_\theta)$, for example, designates the probability measure of $(X_1, \cdots, X_n, Y)$ over $\mathfrak{X}^{n+1}$ where $X_1, \cdots, X_n, Y$ are independent, each X_i has probability measure P_X^θ, and Y has probability measure Q_θ.

For the hypothesis testing problem there may exist a uniformly most powerful test. In this case we would call the corresponding tolerance region most powerful. Failing the existence of a most powerful test, we could look for one yielding a maximum value to the minimum power over the alternative. The corresponding tolerance region we would then call minimax.

6. β-expectation tolerance regions for normal distributions.

6.1. *Univariate normal.* Consider sampling from the univariate normal distribution with density function

$$(2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right],$$

where the parameter space Ω is given by $\mu \,\varepsilon\, R^1$, $\sigma^2 \,\varepsilon\,]0, \infty[$. If a tolerance region is desired which tends to cover the center of the distribution more than the tails, then a reasonable choice of the measure $Q_{\mu\sigma^2}(A)$ on the real line might be the normal probability measure with mean μ and variance $\alpha_1^2\sigma^2$ with $0 < \alpha_1 < 1$. This measure obviously gives more measure to sets in the neighbourhood of μ and less to sets far from μ.

We now consider the analogous hypothesis testing problem. Let $X_1, \cdots, X_n$, Y be independent and let X_i have a normal distribution with mean μ and variance σ^2 and Y have a normal distribution with mean μ, variance $\alpha^2\sigma^2$. The hypothesis testing problem is of the form

$$\text{(6.1.1)}\qquad \begin{array}{lll} \text{Hypothesis:} & \alpha = 1 & (\mu, \sigma^2) \,\varepsilon\, \Omega; \\ \text{Alternative:} & \alpha = \alpha_1 & (\mu, \sigma^2) \,\varepsilon\, \Omega. \end{array}$$

If we define $\bar{x} = n^{-1} \sum x_i$ and $s_x^2 = (n-1)^{-1} \sum (x_i - \bar{x})^2$, then it is easily seen that this problem has the sufficient statistic $(\bar{x}, s_x^2, y)$.

We now apply the invariance method to the problem expressed in terms of the sufficient statistic. Consider the group G of transformations induced by the two groups

$$\text{(6.1.2)}\qquad G_1 = \left\{ \left(\begin{array}{l} \bar{x}' = \bar{x} + a \\ s_x'^2 = s_x^2 \\ y' = y + a \end{array} \right) \middle|\; a \,\varepsilon\, R^1 \right\},$$

and

$$(6.3.1) \qquad G_2 = \left\{ \begin{pmatrix} \bar{x}' = c\bar{x} \\ s_x'^2 = c^2 s_x \\ y' = cy \end{pmatrix} \middle| \; c \,\varepsilon\,]\, 0, \infty\, [\right\}.$$

Obviously, G_1 is a normal subgroup of G. For the group of transformations G_1, a maximal invariant function is $((y - \bar{x}), s_x^2)$. The group G_2 induces a group on this maximal invariant, and it has maximal invariant $\Upsilon = (y - \bar{x})/s_x$. By a theorem of Hunt and Stein [6] Υ is maximal invariant for G.

In accordance with the invariance principle we look for tests based on Υ. Since the variance of $(y - \bar{x})$ is $(\alpha^2 + 1/n)\,\sigma^2$, then under the hypothesis Υ has the distribution of $(1 + 1/n)^{1/2}\, t$, and under the alternative, the distribution of $(\alpha_1^2 + 1/n)^{1/2}\, t$, where t stands for a random variable with Student's t-distribution having $(n - 1)$ degrees of freedom. In terms of Υ, the hypothesis and alternative are simple. To find the most powerful invariant test, we now apply the Neyman-Pearson fundamental lemma. *Let* $c_0 = (1 + 1/n)^{1/2}$, $c_1 = (\alpha_1^2 + 1/n)^{1/2}$ *(clearly* $c_0 > c_1$*). Then, the most powerful test function* $\varphi(\Upsilon)$ *is based on the probability ratio*

$$\frac{\dfrac{1}{c_1[(n-1)\pi]^{1/2}} \dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left[1 + \dfrac{\Upsilon^2}{c_1^2(n-1)}\right]^{-n/2}}{\dfrac{1}{c_0[(n-1)\pi]^{1/2}} \dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left[1 + \dfrac{\Upsilon^2}{c_0^2(n-1)}\right]^{-n/2}}$$

or equivalently on $|\Upsilon|^{-1}$. Hence the most powerful invariant test function is

$$(6.1.4) \qquad \begin{aligned} \varphi(W) &= 1 \quad \text{if } |\Upsilon| < a_\beta \\ &= 0 \quad \text{if } |\Upsilon| > a \end{aligned}$$

and to give the test size β, a_β is $(1 + 1/n)^{1/2}\, t_{1-\beta/2}$, where t_α is the point exceeded with probability α using the Student t-distribution with $(n - 1)$ degrees of freedom.

Since the alternative in (6.1) is a set of the maximal invariant partition of the parameter space, the envelope power function is constant valued over the alternative. By the theorem of Hunt and Stein [5], there is, for any non-invariant test, an invariant test for which the minimum power over the maximal invariant partitions of the parameter space is no smaller. Hence our most powerful invariant test maximizes the minimum power over the alternative among size α tests. Also, it is most stringent.

From the definition of Υ and $\varphi(\Upsilon)$ we have the test

$$\varphi_y(x_1, \cdots, x_n) = 1 \quad \text{if} \left| \frac{y - \bar{x}}{s_x} \right| < a_\beta$$

$$= 0 \quad \text{if} \left| \frac{y - \bar{x}}{s_x} \right| > a_\beta .$$

Thus, the minimax and most stringent tolerance region is

$$S(x_1, \cdots, x_n) = [\bar{x} - a_\beta s_x, \bar{x} + a_\beta s_x].$$

Values of a_β are given in Table 1. It is interesting to note that the tolerance region does not depend on the value of α_1, provided it is less than 1. Also, under the hypothesis, the test statistic has a fixed distribution; hence the test and therefore the tolerance region are similar, and we have a β-expectation similar tolerance region.

If we are interested in having our tolerance region cover the left tail of the distribution, we might choose $Q_{\mu\sigma^2}$ to be the normal distribution with mean $\mu - \epsilon$ and variance $\sigma^2(\epsilon > 0)$. An analysis similar to the above shows that a minimax and most stringent tolerance region is

$$S(x_1, \cdots, x_n) =]-\infty, \bar{x} + a'_\gamma s_x]$$

where a'_γ may be found from Table 1 by using $a'_\gamma = a_{2\gamma-1}$.

If σ^2 is known, our parameter space is given by $\mu \varepsilon R^1$. Using the same Q functions with σ^2 taking the given value, an analysis similar to that above shows that for ability to pick up the center of the distribution, the minimax and most stringent tolerance region is

$$S(x_1, \cdots, x_n) = [\bar{x} - b_\beta \sigma, \bar{x} + b_\beta \sigma]$$

where $b_\beta = (1 + 1/n)^{1/2} z_{(1-\beta)/2}$ and z_α is the point exceeded with probability α using the normal distribution with mean 0 and variance 1. Values of b_β are given in Table 2. Also, the minimax and most stringent tolerance region of size γ which tends to pick up the left-hand tail of the distribution is

$$S(x_1, \cdots, x_n) =]-\infty, \bar{x} + b'_\gamma \sigma]$$

where values of b'_γ may be found from Table 2 by using $b'_\gamma = b_{2\gamma-1}$.

If μ is known, the parameter space is given by $\sigma^2 \varepsilon]0, \infty[$. Using the same Q functions as before with μ taking the given value, a minimax and most stringent size β tolerance region for picking up the center of the distribution is

$$S(x_1, \cdots, x_n) = [\mu - t'_{(1-\beta)/2} s_x, \mu + t'_{(1-\beta)/2} s_x],$$

where s_x is here defined to be $n^{-1} \sum (x_i - \mu)^2$; t'_α is the point exceeded with probability α using Student's t-distribution with n degrees of freedom. Also, the minimax and most stringent size β tolerance region which tends to pick up

TABLE 1

Tolerance factors a_β for univariate normal distributions with unknown mean, unknown variance; sample size n

n	β					
	.995	.99	.975	.95	.90	.75
2	155.9	77.96	31.17	15.56	7.733	2.957
3	16.27	11.46	7.165	4.968	3.372	1.852
4	8.333	6.530	4.669	3.558	2.631	1.591
5	6.132	5.044	3.829	3.041	2.335	1.473
6	5.156	4.355	3.417	2.777	2.176	1.405
7	4.615	3.963	3.174	2.616	2.077	1.361
8	4.274	3.712	3.014	2.508	2.010	1.330
9	4.040	3.537	2.900	2.431	1.960	1.307
10	3.870	3.408	2.816	2.373	1.923	1.290
11	3.741	3.310	2.751	2.327	1.893	1.276
12	3.639	3.233	2.699	2.291	1.869	1.264
13	3.558	3.170	2.657	2.261	1.850	1.255
14	3.491	3.118	2.621	2.236	1.833	1.246
15	3.435	3.074	2.592	2.215	1.819	1.239
16	3.387	3.037	2.567	2.197	1.807	1.234
17	3.346	3.005	2.545	2.181	1.797	1.228
18	3.311	2.978	2.525	2.168	1.787	1.224
19	3.280	2.953	2.509	2.155	1.779	1.220
20	3.252	2.932	2.494	2.145	1.772	1.216
21	3.228	2.912	2.480	2.135	1.765	1.213
22	3.206	2.895	2.468	2.126	1.759	1.210
23	3.186	2.879	2.457	2.119	1.754	1.207
24	3.168	2.865	2.447	2.111	1.749	1.205
25	3.152	2.852	2.438	2.105	1.745	1.202
26	3.137	2.840	2.430	2.099	1.741	1.200
27	3.123	2.830	2.422	2.093	1.737	1.198
28	3.111	2.820	2.415	2.088	1.733	1.197
29	3.099	2.811	2.409	2.083	1.730	1.195
30	3.088	2.802	2.403	2.079	1.727	1.193
31	3.078	2.794	2.397	2.075	1.724	1.192
41	3.007	2.737	2.357	2.046	1.704	1.181
61	2.938	2.682	2.318	2.017	1.684	1.171
121	2.872	2.628	2.279	1.988	1.665	1.161
∞	2.807	2.576	2.241	1.960	1.645	1.150

the left-hand tail is

$$S(x_1, \cdots, x_n) =]-\infty, \mu + t'_{1-\beta} s_x],$$

where t' and s_x are defined immediately above.

6.2 *Multivariate normal.* Consider sampling from a multivariate normal distribution for which the density function is

$$K \exp[-\tfrac{1}{2}(w - \mu)\Lambda(w - \mu)'].$$

TABLE 2

Tolerance factors b_β for univariate normal distributions with unknown mean, known variance; sample size n

n	β					
	.995	.99	.975	.95	.90	.75
2	3.438	3.155	2.745	2.401	2.015	1.409
3	3.241	2.974	2.588	2.263	1.899	1.328
4	3.138	2.880	2.506	2.191	1.839	1.286
5	3.075	2.822	2.455	2.147	1.802	1.260
6	3.032	2.782	2.421	2.117	1.777	1.242
7	3.001	2.754	2.396	2.095	1.758	1.230
8	2.977	2.732	2.377	2.079	1.745	1.220
9	2.959	2.715	2.363	2.066	1.734	1.213
10	2.944	2.702	2.351	2.056	1.725	1.206
11	2.932	2.690	2.341	2.047	1.718	1.201
12	2.922	2.681	2.333	2.040	1.712	1.197
13	2.913	2.673	2.326	2.034	1.707	1.194
14	2.906	2.666	2.320	2.029	1.703	1.191
15	2.899	2.660	2.315	2.024	1.699	1.188
16	2.893	2.655	2.310	2.020	1.696	1.186
17	2.888	2.650	2.306	2.017	1.693	1.184
18	2.884	2.646	2.303	2.014	1.690	1.182
19	2.880	2.643	2.300	2.011	1.688	1.180
20	2.876	2.639	2.297	2.008	1.686	1.179
21	2.873	2.636	2.294	2.006	1.684	1.177
22	2.870	2.634	2.292	2.004	1.682	1.176
23	2.867	2.631	2.290	2.002	1.680	1.175
24	2.865	2.629	2.288	2.000	1.679	1.174
25	2.863	2.627	2.286	1.999	1.677	1.173
26	2.860	2.625	2.284	1.997	1.676	1.172
27	2.859	2.623	2.283	1.996	1.675	1.171
28	2.857	2.621	2.281	1.995	1.674	1.171
29	2.855	2.620	2.280	1.994	1.673	1.170
30	2.853	2.618	2.278	1.992	1.672	1.169
31	2.852	2.617	2.277	1.991	1.671	1.169
41	2.841	2.607	2.269	1.984	1.665	1.164
61	2.830	2.597	2.260	1.976	1.658	1.160
121	2.819	2.586	2.251	1.968	1.652	1.155
∞	2.807	2.576	2.241	1.960	1.645	1.150

Let the parameter space Ω be given by $\mu \varepsilon R^k$ and Λ belonging to the space of $k \times k$ symmetric positive definite matrices. If a tolerance region is wanted which tends to cover the center of the distribution rather than the extremities, then for the parameters μ, Λ a reasonable choice for the $Q_{\mu\Lambda}(A)$ measure over R^k is the normal distribution with mean μ and covariance matrix $\alpha_1^2\Lambda^{-1}$ with $0 < \alpha_1 < 1$.

We now formulate the hypothesis testing problem which corresponds to this

problem in tolerance region construction. Let $W_1, \cdots, W_n, \Xi$ be independent and let each W_i have a normal distribution μ, Λ; and let Ξ have a normal distribution μ, $\alpha^{-2}\Lambda$. Then the problem is to find a best size β test for the problem

$$\text{Hypothesis:} \quad \alpha = 1, \qquad (\mu, \Lambda) \in \Omega,$$

$$\text{Alternative:} \quad \alpha = \alpha_1, \qquad (\mu, \Lambda) \in \Omega.$$

Defining $\bar{w} = n^{-1} \sum_{\alpha=1}^{n} w_\alpha$ and $A = (n-1)^{-1} \sum (w_\alpha - w)' (w_\alpha - w)$, we have as sufficient statistic for this problem, $(\bar{w}, A, \xi)$.

TABLE 3

Tolerance factors c_β for bi-variate normal distributions with unknown means, unknown variance-covariance matrix; sample size n

n	β					
	.995	.99	.975	.95	.90	.75
3	106,667	26,664	4,264	1,064	264.0	40.00
4	746.2	371.2	146.2	71.25	33.75	11.25
5	159.4	98.61	51.34	30.57	17.48	7.295
6	76.66	52.50	31.06	20.25	12.61	5.833
7	50.23	36.41	23.13	15.87	10.37	5.082
8	38.18	28.68	19.06	13.50	9.091	4.626
9	31.50	24.25	16.61	12.03	8.273	4.320
10	27.33	21.41	15.00	11.04	7.705	4.101
11	24.50	19.45	13.85	10.32	7.288	3.936
12	22.47	18.02	13.00	9.778	6.970	3.807
13	20.94	16.93	12.35	9.357	6.719	3.705
14	19.75	16.08	11.83	9.019	6.516	3.620
15	18.81	15.40	11.41	8.743	6.348	3.550
16	18.04	14.83	11.06	8.513	6.208	3.491
17	17.39	14.36	10.76	8.318	6.088	3.440
18	16.85	13.97	10.51	8.151	5.985	3.395
19	16.39	13.62	10.30	8.006	5.895	3.356
20	15.99	13.33	10.11	7.879	5.816	3.322
21	15.64	13.07	9.941	7.768	5.747	3.292
22	15.34	12.84	9.795	7.668	5.685	3.265
23	15.07	12.64	9.663	7.580	5.629	3.240
24	14.82	12.46	9.546	7.500	5.579	3.218
25	14.61	12.29	9.440	7.427	5.533	3.198
26	14.41	12.14	9.343	7.362	5.492	3.179
27	14.23	12.01	9.256	7.302	5.454	3.162
28	14.07	11.89	9.176	7.247	5.419	3.147
29	13.92	11.78	9.102	7.197	5.387	3.133
30	13.79	11.67	9.034	7.150	5.357	3.119
31	13.66	11.58	8.971	7.107	5.330	3.107
32	13.54	11.49	8.913	7.067	5.304	3.095
42	12.73	10.87	8.502	6.783	5.122	3.013
62	11.97	10.28	8.110	6.509	4.945	2.931
122	11.26	9.732	7.736	6.246	4.773	2.851
∞	10.60	9.210	7.378	5.991	4.605	2.773

We apply the invariance method. Consider the group G_1 of transformations on the sample space $R^{k(n+1)}$:

$$G_1 = \left\{ \begin{pmatrix} w'_\alpha = w_\alpha B + \zeta (\alpha = 1, \cdots, n) \\ \xi' = \xi B + \zeta \end{pmatrix} \middle| \begin{array}{l} B \text{ belongs to the class of nonsing-} \\ \text{ular } k \times \text{ matrices, and } \zeta \in R^k \end{array} \right\}.$$

These transformations leave the problem unchanged. The induced group on the

TABLE 4

Tolerance factors c_β for tri-variate normal distributions with unknown means, unknown variance-covariance matrix; sample size n

n	β					
	.995	.99	.975	.95	.90	.75
4	243,169	60,787	9,722	2,427	602.9	92.25
5	1,434	714.0	282.0	138.0	65.96	22.70
6	276.9	171.8	90.06	54.11	31.45	13.74
7	124.8	85.85	51.32	33.90	21.55	10.53
8	78.10	56.98	36.68	25.56	17.10	8.903
9	57.41	43.46	29.33	21.14	14.62	7.931
10	46.17	35.86	24.99	18.44	13.04	7.285
11	39.26	31.05	22.16	16.63	11.96	6.825
12	34.63	27.77	20.17	15.34	11.17	6.481
13	31.33	25.40	18.71	14.38	10.58	6.214
14	28.87	23.62	17.59	13.63	10.11	6.001
15	26.98	22.22	16.70	13.03	9.727	5.827
16	25.47	21.11	15.99	12.54	9.416	5.683
17	24.25	20.20	15.40	12.14	9.156	5.560
18	23.24	19.44	14.90	11.80	8.936	5.456
19	22.39	18.80	14.48	11.51	8.746	5.366
20	21.67	18.25	14.12	11.25	8.581	5.286
21	21.05	17.78	13.81	11.03	8.437	5.216
22	20.51	17.37	13.53	10.84	8.309	5.155
23	20.03	17.00	13.29	10.67	8.196	5.099
24	19.61	16.68	13.07	10.52	8.094	5.049
25	19.24	16.39	12.88	10.38	8.003	5.004
26	18.90	16.14	12.70	10.25	7.920	4.963
27	18.60	15.90	12.54	10.14	7.844	4.926
28	18.33	15.69	12.40	10.04	7.775	4.892
29	18.08	15.50	12.26	9.943	7.712	4.860
30	17.85	15.32	12.14	9.857	7.654	4.831
31	17.64	15.16	12.03	9.777	7.600	4.804
32	17.45	15.01	11.93	9.703	7.550	4.779
33	17.27	14.87	11.83	9.635	7.504	4.756
43	16.04	13.90	11.16	9.150	7.175	4.590
63	14.89	12.99	10.53	8.686	6.857	4.426
123	13.83	12.14	9.922	8.241	6.549	4.266
∞	12.84	11.34	9.348	7.815	6.251	4.108

space of the statistic (w, A, ξ) is

$$G_2 = \left\{ \left(\begin{array}{l} \bar{w}' = \bar{w}B + Z \\ \xi' = \xi B + Z \\ A' = B'AB \end{array} \right) \middle| \begin{array}{l} B \text{ nonsingular,} \\ Z \varepsilon R^k \end{array} \right\}.$$

It is straightforward to show that a maximal invariant statistic is

$$\Upsilon^2 = (\xi - \bar{w})A^{-1}(\xi - \bar{w})'.$$

TABLE 5

Tolerance factors c_β for quadri-variate normal distributions with unknown means, unknown variance-covariance matrix; sample size n

n	β					
	.995	.99	.975	.95	.90	.75
5	432,000	107,992	17,272	4,312	1,072	164.8
6	2,325	1,158	457.9	224.5	107.8	37.71
7	422.4	262.5	138.1	83.36	48.85	21.85
8	182.3	125.8	75.64	50.31	32.34	16.26
9	110.6	81.01	52.54	36.92	25.03	13.46
10	79.38	60.38	41.10	29.92	20.99	11.80
11	62.65	48.91	34.43	25.69	18.46	10.70
12	52.46	41.74	30.11	22.87	16.72	9.916
13	45.70	36.89	27.10	20.87	15.47	9.335
14	40.91	33.40	24.89	19.38	14.52	8.886
15	37.37	30.78	23.22	18.23	13.77	8.528
16	34.64	28.75	21.89	17.31	13.18	8.236
17	32.49	27.13	20.83	16.57	12.69	7.994
18	30.75	25.82	19.95	15.96	12.28	7.790
19	29.32	24.72	19.22	15.44	11.93	7.615
20	28.12	23.83	18.60	15.00	11.63	7.464
21	27.10	23.02	18.07	14.62	11.38	7.332
22	26.22	22.34	17.60	14.28	11.15	7.216
23	25.46	21.75	17.20	13.99	10.95	7.113
24	24.79	21.23	16.84	13.73	10.78	7.021
25	24.20	20.77	16.52	13.50	10.62	6.938
26	23.68	20.36	16.24	13.30	10.48	6.863
27	23.21	19.99	15.98	13.11	10.35	6.795
28	22.79	19.66	15.75	12.94	10.23	6.733
29	22.41	19.36	15.54	12.79	10.12	6.676
30	22.06	19.09	15.35	12.64	10.03	6.624
31	21.74	18.84	15.17	12.51	9.935	6.576
32	21.45	18.61	15.01	12.40	9.851	6.532
33	21.19	18.39	14.86	12.28	9.774	6.490
34	20.94	18.20	14.72	12.18	9.703	6.452
44	19.23	16.84	13.75	11.46	9.195	6.177
64	17.66	15.57	12.83	10.77	8.706	5.907
124	16.20	14.38	11.96	10.11	8.234	5.643
∞	14.86	13.28	11.14	9.488	7.780	5.385

The problem as interpreted for the induced distributions of the statistic Υ has a simple hypothesis and a simple alternative. By applying the Neyman-Pearson fundamental lemma, a short analysis shows that the most powerful invariant test is

$$\begin{aligned}\varphi(\Upsilon^2) &= 1 \qquad \text{if } \Upsilon^2 < c_\beta \\ &= 0 \qquad \text{if } \Upsilon^2 > c_\beta .\end{aligned}$$

Under the hypothesis, $\Upsilon^2\ (1 + 1/n)^{-1}$ has the distribution of Hotelling's T^2 with $(n - 1)$ degrees of freedom. The probability density function of T^2 with $(n - 1)$ degrees of freedom is

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)} \frac{(T^2/n-1)^{(k-2)/2}}{(1 + T^2/n-1)^{n/2}}\, d(T^2/n - 1). \tag{6.2.1}$$

But if we make the transformation $T^2 = (n - 1)\ k/(n - k)\ F$, (6.2.1) is easily seen to become the probability density function of Fisher's F-distribution with $k, n - k$ degrees of freedom. Hence, to give the test and consequently the tolerance region size β, we take

$$c_\beta = (1 + 1/n)\ (n - 1)\ (k/n - k)\ F_{1-\beta} ,$$

where F_α is the point exceeded with probability α using the F-distribution with $k, n - k$ degrees of freedom.

Now, by the same argument used for the univariate case, the minimax and most stringent size β tolerance region for the k-variate normal distribution is the ellipsoidal region given by

$$\{\xi \mid (\xi - \varpi)\ A^{-1}\ (\xi - \varpi)' \leqq c_\beta\}.$$

Values of c_β for $k = 2, 3, 4$ are given in Tables 3, 4, and 5.

REFERENCES

[1] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin & Company, Ltd., London, 1948.

[2] P. Halmos, "The theory of unbiased estimation," *Ann. Math. Stat.*, Vol. 17 (1946), p. 34–39.

[3] E. L. Lehmann, "Notes on the theory of estimation," University of California Press, 1950.

[4] D. A. S. Fraser, "Completeness of order statistics," *Canadian J. Math.*, Vol. 6 (1953), pp. 42–45.

[5] E. L. Lehmann, "Some principles of the theory of hypothesis testing," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.

[6] E. L. Lehmann, "Theory of testing hypothesis," lecture notes published at the University of California, 1949.

GENERALIZED TOLERANCE LIMITS

By J. H. B. Kemperman

Purdue University

1. Summary. A method for constructing tolerance limits due to Fraser [8] is generalized by allowing that each step of the construction may depend not only on the blocks previously formed but also on all the known boundary observations and, moreover, on certain sets of indices. Furthermore, Tukey's [5] lexicographical ordering is replaced by a more general type of ordering.

2. Introduction. Let $\{\Omega, \mathfrak{A}, \mu(A)\}$ be a measure space with $\mu(\Omega) = 1$ and $\mathfrak{A}$ complete. Then the relation $P(X \varepsilon A) = \mu(A)$ $(A \varepsilon \mathfrak{A})$ defines a random variable X taking values in the space Ω. Let $W = (x_1, \cdots, x_n)$ be a set of n independent observations on X and let $D_j = D_j(W)$ $(j = 1, 2, \cdots)$ be disjoint measurable subsets of Ω depending on W. These D_j sets are called (nonparametric) tolerance limits when the joint distribution of the random "coverages" $\mu(D_j)$ does not depend on the true distribution $\mu(A)$ of X, given that the latter belongs to a certain rather wide class of probability measures. Such tolerance limits were first introduced by S. S. Wilks [1] whose method was generalized to a far extent by A. Wald, H. Scheffé, J. W. Tukey, R. Wormleighton, and D. A. S. Fraser ([2], [3], [4], [5], [6], [7], [8]).

3. Ordering. By a (generalized) ordering o in Ω we shall mean an assignment of exactly one of the relations $x_1 < x_2$, $x_1 \sim x_2$, or $x_1 > x_2$ to each pair x_1, x_2 of points in Ω, such that $x_1 \sim x_2$ is an equivalence relation and such that o induces an (ordinary) transitive ordering among the corresponding equivalence classes. Let $\Omega = A \cup B$ with $A < B$ in the obvious sense. We shall assume that always: (i) A is measurable. (ii) If A is non-empty, we have $A = \bigcup_k \{x \mid x \leqq a_k\}$ for some (at most denumerable) subsequence $\{a_k\}$ of A. Similarly, if B is non-empty, we have $A = \bigcap_k \{x \mid x < b_k\}$ for some subsequence $\{b_k\}$ of B.

One way of obtaining such a generalized ordering is as follows: Let M be a finite or denumerable well-ordered set and let, for each m in M, $g_m(x)$ be a real-valued measurable function on Ω. If $g_m(x_1) = g_m(x_2)$ for all m in M, we define $x_1 \sim x_2$. Otherwise, $x_1 < x_2$ if and only if $g_s(x_1) < g_s(x_2)$, where s is the smallest index such that $g_s(x_1) \neq g_s(x_2)$.

An ordering o is said to be *continuous* (with respect to the measure $\mu(A)$) when for each x_0 in Ω we have $\mu\{x \mid x \sim x_0\} = 0$.

Lemma 1. *Let o be a continuous ordering and let $q(x_0) = P(X < x_0) = \mu\{x \mid x < x_0\}$. Then $q(X)$ is a uniformly distributed random variable in* [0, 1].

Proof. Let $0 \leqq q \leqq 1$, $A = \{x \mid q(x) \leqq q\}$, and $B = \{x \mid q(x) > q\}$. We have to show that

$$P(q(X) \leqq q) = P(X \varepsilon A) = \mu(A) = q.$$

Received September 11, 1954, revised January 29, 1955.

But for a_k in A, we have $\mu\{x \mid x \leqq a_k\} = \mu\{x \mid x < a_k\} = q(a_k) \leqq q$; hence, from (ii), $\mu(A) \leqq q$ whether or not A is empty. Moreover, for b_k in B, we have $\mu\{x \mid x < b_k\} = q(b_k) > q$; hence, from (ii), $\mu(A) \geqq q$ whether or not B is empty.

4. Partitioning. Let m, m_0, and m_1 be positive integers, $m = m_0 + m_1$. Let $x_1, \cdots, x_{m-1}$ be $m - 1$ points in a measurable subset D of Ω and let o be a given ordering. Denoting by x^* the m_0-th smallest ($= m_1$-th largest) point x_i with respect to o, the partition of D into the three disjoint subsets $D_0 = \{x \mid x < x^*, x \varepsilon D\}$, $D_1 = \{x \mid x > x^*, x \varepsilon D\}$, and $D^* = \{x \mid x \sim x^*, x \varepsilon D\}$ is called the (m_0, m_1)-partition of D with respect to o and to the $m - 1$ points x_i in D. Note that, when $x_i \sim x_j$ does not happen for $i \neq j$, the "boundary" element x^* is unique, while D_0, D_1, and D^* contain exactly $m_0 - 1$, $m_1 - 1$, and 1 elements x_i, respectively. If o is continuous, $\mu(D^*) = 0$, hence, $\mu(D) = \mu(D_0) + \mu(D_1)$.

LEMMA 2. *If $\mu(D) > 0$ we assume that o is continuous and that $x_1, \cdots, x_{m-1}$ are $m - 1$ independent observations on X restricted to $X \varepsilon D$. Then $\mu(D_0) = q\mu(D)$, where q is a random variable which has the incomplete Beta-function $I_q(m_0, m_1)$ as its cumulative distribution function.*

PROOF. We may assume that $\mu(D) > 0$. Let Y be the random variable whose distribution $\nu(A) = \mu(A)\mu(D)^{-1}(A \subset D)$ is that of X restricted to $X \varepsilon D$. Observing that o induces an ordering on D which is continuous with respect to $\nu(A)$, it follows from Lemma 1 (replacing Ω by D, and X by Y) that for $q(x_0) = \nu\{x \mid x < x_0, x \varepsilon D\}$ the variable $q(Y)$ is uniformly distributed in [0, 1]. Hence, $q(x^*)$ is the m_0-th smallest among $m - 1 = m_0 + m_1 - 1$ independent observations $q(x_i)$ on a uniformly distributed random variable in [0, 1]. This proves that $q(x^*) = \mu(D_0)\mu(D)^{-1}$ has the d.f. $I_q(m_0, m_1)$.

5. The construction. For the sake of brevity, we shall employ a somewhat colloquial language. In the construction two persons are involved: a statistician (S) and his assistant (A). A knows precisely the actual outcomes of the n independent observations $x_1, \cdots, x_n$ on X, while, at the very outset, S has no information at all about these outcomes. On the other hand, S has at his disposal a class H of orderings o in Ω known to be continuous with respect to the distribution $\mu(A)$ of X.

In the first step of the construction, S selects an ordering 0_1 from H and a positive integer m_0, $m_0 \leqq n$, and asks A to give him the m_0-th smallest observation $x^*(1)$ with respect to 0_1 (this element is unique with probability 1), together with the two sets of indices corresponding to the $m_0 - 1$ and $m_1 - 1 = n - m_0$ observations which are smaller or larger than $x^*(1)$, respectively. Now, S can draw the (m_0, m_1)-partition $\Omega = \Omega_0 \cup \Omega^* \cup \Omega_1$ of Ω with respect to 0_1 and the set of the n observations x_i in Ω. Let $D^0(0) = \Omega$, $D^1(j) = \Omega_j$ ($j = 0, 1$), and $D_1^* = \Omega^*$.

After k steps, $0 \leqq k \leqq n - 1$, S has obtained a partition of Ω into $k + 1$ disjoint "blocks" $D^k(j)$ ($j = 0, 1, \cdots, k$) and k boundary sets D_i^* ($i = 1, \cdots, k$), each of μ-measure 0. Further, for each of these $2k + 1$ sets, S knows precisely the set of indices corresponding to the observations x_i within the set.

Finally, for each boundary set D_i^* $(i = 1, \cdots, k)$, S knows the actual value of the boundary observation $x^*(i)$ in D_i^* (with a probability 1 these boundary observations are unique).

Now, the $(k + 1)$-th step of the construction proceeds as follows: His choice depending, *in any way whatsoever*,[1] on the knowledge acquired, S chooses: (i) A distinguished block $D = D^k(j^*)$ among those of the $k + 1$ blocks $D^k(j)$ $(j = 0, \cdots, k)$ which contain at least one observation. (ii) A positive integer m_0 not larger than the number $m - 1$ of observations in D. (iii) An ordering 0_{k+1} from H.

He then asks A for the m_0-th smallest observation $x^*(k + 1)$ in D with respect to 0_{k+1}, together with the two sets of indices corresponding to the $m_0 - 1$ or $m_1 - 1 = m - m_0 - 1$ observations in D which are smaller or larger than $x^*(k + 1)$, respectively.

Using the acquired value $x^*(k + 1)$, S is now able to draw the (m_0, m_1)-partition $D = D_0 \cup D^* \cup D_1$ of D with respect to 0_{k+1} and the $m - 1$ observations in D. Afterwards, he renumbers the blocks $D^k(0), \cdots, D^k(j^* - 1), D_0, D_1, D^k(j^* + 1), \cdots, D^k(k)$ as $D^{k+1}(j)$ $(j = 0, \cdots, k + 1)$, in this order. Finally, let $D_{k+1}^* = D^*$.

After exactly n steps the construction stops. Then S has obtained a partition of Ω into $n + 1$ disjoint blocks $U_j = D^n(j)$ $(j = 0, \cdots, n)$ and, further, n boundary sets D_k^* $(k = 1, \cdots, n)$, each of μ-measure 0.

THEOREM 1. *The coverages* $c_j = \mu(U_j)$ $(j = 0, \cdots, n)$ *have the joint distribution* $dc_1\, dc_2 \cdots dc_n$, *where* $c_j \geqq 0$, $c_1 + \cdots + c_n = 1 - c_0 \leqq 1$. *Moreover, the union U of m distinct sets* U_j *has a coverage* $p = \mu(U)$ *with* d.f. $I_p(m, n + 1 - m)$.

Let $0 < \alpha < 1$, and let $p = p_m(\alpha)$ be such that $I_p(m, n + 1 - m) = \alpha$. Then with a probability $1 - \alpha$, the random set U contains at least a proportion $p_m(\alpha)$ of the total probability mass 1 in Ω (i.e., we have confidence limits on the distribution of X or its parameters). For $\alpha = .01$ or $.05$, the value $p_m(\alpha)$ may be determined by using F-tables. Let F_0 be the α-point of the F-distribution with $n_1 = 2(n + 1 - m)$ and $n_2 = 2m$ degrees of freedom. Then $p_m(\alpha) = (1 + F_0 n_1/n_2)^{-1}$.

Some warning seems desirable. If the construction stops after k steps, we have a partition of Ω into the blocks $D^k(j)$ $(j = 0, \cdots, k)$ and, further, k boundary sets of measure 0. Let the random variable $n_j - 1$ denote the number of observations in $D^k(j)$ $(j = 0, \cdots, k)$, and let $N_j = n_0 + \cdots + n_{j-1}$. One can easily see that $D^k(j)$ is the union of the "final" blocks $U_{N_j}, \cdots, U_{N_j+n_j-1}$ (which might be found by completing the construction) and, further, some set of measure 0. However (as certain counterexamples show), this does not imply that conditional to $n_j = m$ (m given) the coverage of $D^k(j)$ has the conditional dis-

[1] Chance decisions are also allowed. For example, instead of making each decision as the necessity for it arises, S could start with a complete plan which provides for all contingencies. Then we may as well assume that S has already determined beforehand the actual outcomes of the random decisions which might arise.

tribution $I_y(m, n + 1 - m)$. Generally, the latter conclusion is only justified when (with a probability 1) both N_j and n_j are constant.

6. Proof of Theorem 1. In order to keep the proof on an elementary level, we shall avoid an explicit use of the usual complicated measure preserving transformations (cf. Fraser [8], pp. 53–54). Let $t_j = t(j) = s_j\sqrt{-1}$ $(j = 0, \cdots, n)$ be complex parameters with s_j real. It suffices to show that the characteristic function

$$E[\exp(t_0 \log c_0 + \cdots + t_n \log c_n)] = E(c_0^{t_0} \cdots c_n^{t_n})$$

depends only on n and the t_j but not on the distribution of X or the actual mode of construction. For then the joint distribution of $\log c_0, \cdots, \log c_n$, and hence the joint distribution of $c_0, \cdots, c_n$, will not be affected when X is replaced by a real random variable, uniformly distributed in $[0, 1]$, and when the ordering 0_k $(k = 1, \cdots, n)$ is replaced by the common ordering in $[0, 1]$. But then $c_0, \cdots, c_n$ become the differences between consecutive order statistics, and Theorem 1 now follows from a well-known result (cf. Wilks [1]).

After k steps, $0 \leqq k \leqq n$, the construction based on the sample $W = (x_1, \cdots, x_n)$, yields (with a probability 1) a partition of Ω into the $k + 1$ blocks $D^k(j)$ $(j = 0, \cdots, n)$ and, further, k boundary sets of measure 0. Let the random variable $n_j - 1$ equal the number of observations inside $D^k(j)$ and let $N_j = n_0 + n_1 + \cdots + n_{j-1}$. Now consider the quantity

$$\rho_k = \prod_{j=0}^{k} \frac{\Gamma(n_j)\, \mu(D^k(j))^{t(N_j)+\cdots+t(N_j+n_j-1)}}{\Gamma[n_j + t(N_j) + \cdots + t(N_j + n_j - 1)]},$$

depending on the parameters $t_0, \cdots, t_n$. Though for $1 < k < n$ the joint distribution of $\mu(D^k(0)), \cdots, \mu(D^k(k))$ depends strongly on S's method of construction, it turns out that, for any mode of construction and for each (arbitrary but fixed) set of values $t_0, \cdots, t_n$,

$$E(\rho_k) = n!\, \Gamma(n + 1 + t_0 + \cdots + t_n)^{-1} \qquad (k = 0, 1, \cdots, n). \tag{1}$$

For $k = n$, we have $n_j = 1, N_j = j, \mu(D^k(j)) = \mu(U_j) = c_j$ $(j = 0, \cdots, n)$, and (1) implies

$$E(c_0^{t_0} \cdots c_n^{t_n}) = n!\, \Gamma(n + 1 + t_0 + \cdots + t_n)^{-1} \prod_{j=0}^{n} \Gamma(t_j + 1),$$

where indeed the right-hand side depends only on n and the t_j.

Formula (1) is evident for $k = 0$. For, $\mu(D^0(0)) = \mu(\Omega) = 1$ and $n_0 = n + 1$, $N_0 = 0$ (when $k = 0$) imply that ρ_0 is always equal to the right-hand side of (1). Let k be a fixed integer, $0 \leqq k \leqq n - 1$; it suffices to prove that $E(\rho_k) = E(\rho_{k+1})$.

Let $D^k(j)$ $(j = 0, \cdots, k)$, $D = D^k(j^*)$, D_0, D_1, m, m_0, and m_1 be as defined in the $(k + 1)$-th step of the construction. Here, with probability 1, D_0 and D_1 contain precisely $m_0 - 1$ and $m_1 - 1$ observations, respectively; $(m_0 + m_1 = m)$. Moreover, $\mu(D) = \mu(D_0) + \mu(D_1)$. It now follows from the definitions of ρ_k,

ρ_{k+1}, and the blocks $D^{k+1}(j)$ $(j = 0, \cdots, k+1)$ that

$$\rho_{k+1} = \rho_k \frac{\Gamma(m_0)\,\Gamma(m_1)\,\Gamma(m + t' + t'')}{\Gamma(m)\,\Gamma(m_0 + t')\,\Gamma(m_1 + t'')} q^{t'}(1 - q)^{t''}, \tag{2}$$

where $\mu(D_0) = q\mu(D)$ and, with $N = N_{j^*}$,

$$t' = t(N) + \cdots + t(N + m_0 - 1), \qquad t'' = t(N + m_0) + \cdots + t(N + m - 1).$$

In view of the footnote to the construction, we may assume without loss of generality that S has a complete non-random plan of construction which provides for all contingencies. The following information Σ has been received by S from A during the first k steps of the construction: (i) For $i = 1, \cdots, k$, the value ξ_i and the index ν_i of the boundary observation $x^*(i)$. (ii) For $j = 0, \cdots, k$, the indices $\sigma(j, h)$ $(h = 1, \cdots, n_j - 1)$ of the observations in the block $D^k(j)$. Here, the n different integers ν_i and $\sigma(j, h)$ together constitute the full set of indices $(1, 2, \cdots, n)$.

Knowing only Σ, S can reconstruct the blocks $D^k(j)$ $(j = 0, \cdots, k)$ according to plan; hence, Σ is equivalent to the information known to S at the beginning of the $(k + 1)$-th step. Therefore, Σ completely determines the quantity ρ_k, the distinguished block $D = D^k(j^*)$, together with the ordering 0_{k+1}, and the positive integers m_0 and m_1 $(m_0 + m_1 = m = n_{j^*})$ mentioned in the $(k + 1)$-th step of the construction.

To almost all samples W there corresponds a set of information Σ of the above type. Among these corresponding Σ's, let Σ_0 be a specific set of information (i) and (ii). Denoting the ith observation by $x(i)$, it is evident that in an actual construction Σ_0 will arise if and only if: (i) $x(\nu_i) = \xi_i$ $(i = 1, \cdots, k)$. (ii) For $j = 0, \cdots, k$, we have $x(\sigma(j, h)) \in D^k(j)$ $(j = 1, \cdots, n_j - 1)$, where the set $D^k(j)$ is uniquely determined by Σ_0. Hence, we have for $0 \leqq j \leqq k$ that, given $\Sigma = \Sigma_0$, the observations $x(\sigma(j, h))$ $(h = 1, \cdots, n_j - 1)$ behave as $n_j - 1$ independent observations on the random variable X restricted to $X \in D^k(j)$ (provided $\mu(D^k(j)) > 0$).

Further, D_0 is obtained as the "lower" set in the (m_0, m_1)-partition of $D = D^k(j^*)$ with respect to the $n_{j^*} - 1 = m - 1$ observations $x(\sigma(j^*, h))$ in D and the continuous ordering 0_{k+1}. It follows from Lemma 2 that, given $\Sigma = \Sigma_0$, we have $\mu(D_0) = q\mu(D)$, where q has the conditional d.f. $I_q(m_0, m_1)$.

Moreover, given $\Sigma = \Sigma_0$, the quantities ρ_k, m, m_0, m_1, t', and t'' are constants. It now follows from (2) that

$$E(\rho_{k+1} \mid \Sigma = \Sigma_0) = \rho_k = E(\rho_k \mid \Sigma = \Sigma_0),$$

implying that $E(\rho_{k+1}) = E(\rho_k)$.

7. A remark. The above proof is not completely rigorous because the very last step ("implying that") is still open to doubt for lack of a precise definition of the expected values $E(\rho_{k+1} \mid \Sigma = \Sigma_0)$, $E(\rho_{k+1})$, etc. The latter omission is also the root of the following difficulty.

If, in the construction, S's decisions depend too wildly (that is, in a non-measurable way) on the available information, it may easily happen that the coverage c_j of the final block U_j is a non-measurable function (with respect to the Borel field $\mathfrak{A}^n$ in Ω^n) of the sample point $W = (x_1, \cdots, x_n)$. Then the question arises as to what (in the assertion of Theorem 1) is meant by the probability $\Pr(c_j \leqq a_j)$ $(j = 0, \cdots, n)$. The following approach to this question, which avoids additional measurability assumptions, was indicated to me by Prof. D. A. S. Fraser.

For simplicity, let us assume that S starts with a complete *non-random* plan which provides for all contingencies. Let Q stand for a specific (a priori possible) outcome of the indices of the observations inside the blocks $D^k(j)$ $(k = 1, \cdots, n;$ $j = 0, \cdots, k)$ and the indices i_k of the boundary observations $x^*(k)$ $(k = 1,$ $\cdots, n)$. Let the (finitely many) different possible outcomes Q be denoted by $Q_1, \cdots, Q_p$. Let $f_r(W) = 1$ $(r = 1, \cdots, p)$ when the construction based on W yields the outcome Q_r; otherwise, $f_r(W) = 0$. Thus, $\sum_r f_r(W) = 1$ for almost all W.

Let F_n be the class of all the subsets B of Ω^n such that, for $r = 1, \cdots, p$, the integral

$$P_r(B) = \int_{\Omega^n} [[f_r(W)\, \chi_B(W)\, d\mu(x_{i_n})] \cdots d\mu(x_{i_1})]$$

has a meaning and exists as a *repeated* Lebesgue-Stieltjes integral; here $(i_1, \cdots,$ $i_n)$ corresponds to Q_r, while $\chi_B(W)$ denotes the characteristic function of B. Let F_k $(0 \leqq k < n)$ be the class of F_n-sets B such that $W_1 \subset B$ implies $W_2 \subset B$ whenever the two constructions based on W_1 and W_2 yield, at the end of the kth step, exactly the same information Σ (cf. the above proof).

One can show that: (1) F_k is a Borel field $(k = 0, \cdots, n)$ and $F_0 \subset F_1 \subset \cdots \subset$ F_n. (ii) $P(B) = \sum_r P_r(B)$ defines a probability measure on F_n. (iii) The function $\rho_k(W)$, employed in the above proof, is F_k-measurable $(k = 0, \cdots, n)$; hence, $c_j = c_j(W)$ is F_n-measurable. (iv) The above proof becomes exact by defining (at the $(k + 1)$-th step) $E(y \mid \Sigma = \Sigma_0)$ as the conditional expectation of y relative to F_k with $\{F_n, P(B)\}$ as the underlying measure space. (v) Consequently, interpreting the assertion of Theorem 1 in terms of this same measure space, we have a meaningful and true result.

8. The discontinuous case. The above procedure imposes one restriction on the distribution $\mu(A)$ of X; namely, that each ordering (which might be used in the construction) of the given class H is a continuous ordering with respect to $\mu(A)$. In the so-called discontinuous case, the distribution $\mu(A)$ of X is completely unrestricted. However, in this case the above construction might break down with a positive probability in the sense that some boundary set will contain more than one observation. This defect will be repaired as follows (cf. Fraser [8], p. 50).

Let Y be a real random variable, uniformly distributed in the unit interval

$L = [0, 1]$, which is independent of X and let $X' = (X, Y)$, taking values in $\Omega' = \Omega \times L$. To each ordering o in Ω we associate the following ordering o' in Ω':

$$(x_1, y_1) < (x_2, y_2) \qquad \text{if } x_1 < x_2 \text{ or } x_1 \sim x_2 \text{ and } y_1 < y_2 .$$

Let H' consist of all orderings in Ω' which are associated to some ordering in Ω. Then, even in the discontinuous case, each ordering o' in H' is continuous with respect to the distribution $\mu'(B)$ of X'.

Let $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ be independent observations on X and Y, respectively. Then $x_i' = (x_i, y_i)$ $(i = 1, \cdots, n)$ are n independent observations on X'. Replacing in the above construction Ω, H, and x_i by Ω', H', and x_i', respectively, we obtain a partition of Ω' into the final blocks U_j' $(j = 0, \cdots, n)$ and the set of measure 0 consisting of the n observations x_i'. Clearly, the coverages $c_j = \mu'(U_j')$ satisfy the assertions of Theorem 1. Thus we are able to set precise tolerance limits on the distribution of $X' = (X, Y)$ which will yield some information on the distribution of X.

As a simple illustration: Let o be any ordering in Ω and let $x'(1) \leqq x'(2) \leqq \cdots \leqq x'(n)$ be the ordered set (with respect to o') of the n observations x_i' on X'. Then $U = \{x' \mid x' < x'(m)\}$ has a coverage $p = \mu'(U)$ with a cumulative d.f. $I_p(m, n + 1 - m)$. But, for $x'(m) = (x(m), y(m))$,

$$\begin{aligned} \mu'(U) &= \mu\{x \mid x < x(m)\} + y(m)\mu\{x \mid x \sim x(m)\} \\ &\geqq \mu\{x \mid x < x(m)\} = P(X < x(m)) = c \qquad \text{(say)}. \end{aligned}$$

Hence,

$$P(c \leqq p) \geqq P(\mu'(U) \leqq p) = I_p(m, n + 1 - m),$$

a well-known result due to Scheffé and Tukey ([3], p. 191).

REFERENCES

[1] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 91–96.

[2] A. Wald, "An extension of Wilks' method for setting tolerance limits," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[3] H. Scheffé and J. W. Tukey, "Non-parametric estimation: I. Validation of order statistics," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187–192.

[4] J. W. Tukey, "Nonparametric estimation: II. Statistically equivalent blocks and tolerance regions—the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.

[5] J. W. Tukey, "Nonparametric estimation: III. Statistically equivalent blocks and multivariate tolerance regions—the discontinuous case," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 30–39.

[6] D. A. S. Fraser and R. Wormleighton, Nonparametric estimation: IV.," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 294–298.

[7] D. A. S. Fraser, "Sequentially determined statistically equivalent blocks," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 372–381.

[8] D. A. S. Fraser, "Nonparametric tolerance regions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 44–55.

ON A CHARACTERIZATION OF THE STABLE LAW WITH FINITE EXPECTATION

BY R. G. LAHA

Indian Statistical Institute

1. Introduction and summary. A remarkable characterization of the normal law is that if x and y are two independent chance variables such that two linear functions, $ax + by$ $(ab \neq 0)$ and $cx + dy$ $(cd \neq 0)$, are distributed independently of each other, then both x and y are normally distributed. This theorem has been proved without any assumption about the existence of moments by Darmois [2], extending earlier results of Gnedenko [4] and Kac [5]. The question that naturally arises is how far the condition of stochastic independence is necessary, or, in other words, whether the above theorem can be generalised after relaxing the condition of stochastic independence of the linear functions of two independent chance variables. But it is evident that we can always construct two linear functions of non-normal mutually independent chance variables such that they are not independent in the probability sense. In the present paper we shall investigate the nature of the distribution law that may be obtained by imposing the mild restriction of the linearity of regression of one linear function on the other, which is, of course, weaker than the assumption of stochastic independence. We shall prove a general theorem from which a number of results will follow as special cases. But it should be noted that the statements regarding regression or conditional expectation require the assumption that the conditional distribution function exist, and in the following, this assumption will be tacitly made wherever needed.

2. Results. First of all, we shall give a short proof of the following lemma of Rao [8], Rothschild and Mourier [10].

LEMMA. *Let x and y be two proper random variables each having a finite expectation (which we may assume to be zero without any loss of generality) such that the regression of y on x exists. Then the necessary and sufficient condition for the regression of y on x to be linear is that*

$$\left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} = \beta \frac{d\varphi(u, 0)}{du},$$

where $\varphi(u, v)$ stands for the characteristic function of the joint cumulative distribution of x and y, and β is a constant.

PROOF OF NECESSITY. Since $\varphi(u, v)$ represents the characteristic function of

Received October 11, 1954; revised March 17, 1955.

the joint cumulative distribution of x and y, we have

$$\begin{aligned}\varphi(u, v) &= E(e^{iux+ivy}) \\ &= \int e^{iux+ivy}\, dF(x, y) \\ &= \int e^{iux}\left[\int e^{ivy}\, dF_x(y)\right] dF(x),\end{aligned}$$

where $F_x(y)$ represents the conditional distribution function of y for fixed x.

But since the expectations of both x and y are assumed to exist and to be equal to zero and, further, the regression of y on x is linear, we must have

$$\begin{aligned}\left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} &= i\int e^{iux}\left[\int y\, dF_x(y)\right] dF(x) \\ &= i\beta\int e^{iux}x\, dF(x) \\ &= \beta\frac{d\varphi(u, 0)}{du}.\end{aligned}$$

Proof of Sufficiency. Since the regression of y on x is assumed to exist, let us denote it by $E_x(y)$, so that $E_x(y) = \int y\, dF_x(y)$.

Then proceeding as above, it can be very easily shown that

$$\left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} = i\int e^{iux}[E_x(y)]\, dF(x).$$

Hence, the condition

$$\left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} = \beta\frac{d\varphi(u, 0)}{du}$$

gives

$$\int e^{iux}[E_x(y) - \beta x]\, dF(x) = 0.$$

Then, from the uniqueness theorem of Fourier transforms of functions of bounded variation, it follows that $E_x(y) \equiv \beta x$, for all x, except for a set of probability measure zero.

Theorem 1. *Let x, ξ, and η be three proper random variables, each having a finite expectation (which may be assumed to be zero without any loss of generality) such that x is distributed independently of the joint distribution of ξ and η, but ξ and η have a joint distribution where the regression of η on ξ exists and is linear and given by $E_\xi(\eta) = \beta_0\xi$. Then the regression of $Y = cx + \eta$ on $X = ax + \xi$, $(a \neq 0)$, is always linear irrespective of the distribution functions of x, ξ, and η, whenever the relationship $c = a\beta_0$ is satisfied.*

Proof. Let $\Phi(u, v)$, $\varphi(u, v)$, and $\varphi_1(u)$ represent the characteristic functions

of the joint cumulative distribution of (X, Y), (ξ, η), and the cumulative distribution of x, respectively. Then,

$$\Phi(u, v) = E\{e^{iuX+ivY}\}$$

$$(1) \qquad = E\{e^{iu(ax+\xi)+iv(cx+\eta)}\}$$

$$= \varphi_1(au + cv)\varphi(u, v).$$

Now, differentiating both sides of (1) with respect to v and then putting $v = 0$, we get

$$(2) \qquad \left.\frac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = c\varphi_1'(au)\varphi(u, 0) + \left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} \varphi_1(au).$$

But, using the lemma above, since $E_\xi(\eta) = \beta_0\xi$,

$$(3) \qquad \left.\frac{\partial\varphi(u, v)}{\partial v}\right]_{v=0} = \beta_0 \frac{d\varphi(u, 0)}{du}.$$

Next, substituting (3) in (2), we get

$$(4) \qquad \left.\frac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = c\varphi_1'(au)\varphi(u, 0) + \beta_0\varphi'(u, 0)\varphi_1(au).$$

Again, putting $v = 0$ in (1) and then differentiating both sides with respect to u, we get

$$(5) \qquad \frac{d\Phi(u, 0)}{du} = a\varphi_1'(au)\varphi(u, 0) + \varphi'(u, 0)\varphi_1(au).$$

Now, if $c = a\beta_0$, substituting this value of c in (4) and then comparing with (5), we get easily

$$(6) \qquad \left.\frac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = \beta_0 \frac{d\Phi(u, 0)}{du}.$$

Then from the lemma above, it follows that the regression of Y on X is always linear, whatever may be the distribution function of x, ξ, and η. From Theorem 1, it follows that if η and ξ are stochastically independent, and further if $c = 0$, then the regression of Y on X is always linear, since in this case $\beta_0 = 0$ and the relationship $c = a\beta_0$ is satisfied.

Similarly, if $\xi = by$ $(b \neq 0)$ and $\eta = dy$ $(d \neq 0)$, and further if $bc = ad$, the regression of Y on X is always linear.

3. Further results.

THEOREM 2. *With the same notations and assumptions as used in Theorem* 1, *the necessary and sufficient condition for the regression of Y on X to be linear for all a contained in a closed interval (a_1, a_2), where either $a_1 < a_2 < 0$ or $0 < a_1 < a_2$, and for some c for which the relationship $c \neq a\beta_0$ is satisfied for all a in the interval, is that both x and ξ should belong to a class of stable law with finite expectation.*

Proof of Necessity. Using the above lemma, the condition of the linearity of regression Y on X gives the relation,

$$\left.\frac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = \beta\frac{d\Phi(u, 0)}{du}. \tag{7}$$

Next, using (4), (5), and (7) together, we get, after a little rearrangement of terms,

$$(c - a\beta)\varphi_1'(au)\varphi(u, 0) = (\beta - \beta_0)\varphi_1(au)\varphi'(u, 0). \tag{8}$$

We shall first show that in (8) neither $c - a\beta$ nor $\beta - \beta_0$ can be equal to zero under the conditions of the theorem.

Let us suppose that $\beta - \beta_0 = 0$ when $c - a\beta \neq 0$. In this case, (8) reduces to

$$\varphi_1'(au)\varphi(u, 0) = 0. \tag{9}$$

Since $\varphi(u, 0)$ is continuous and equal to unity at the origin, $u = 0$, there always exists a neighbourhood, say $u_0 > 0$, such that for all $|u| < u_0$, we have $\varphi(u, 0) \neq 0$. Then it follows from (9) that for $|u| < u_0/a_0$, we have

$$\varphi_1'(u) = 0, \tag{10}$$

where a_0 is the larger of $|a_1|$ and $|a_2|$. This implies that the distribution of x itself is improper, the whole mass being concentrated at the origin $x = 0$.

Similarly, if $c - a\beta = 0$ when $\beta - \beta_0 \neq 0$, (8) reduces to

$$\varphi_1(au)\varphi'(u, 0) = 0. \tag{11}$$

From (11), proceeding exactly as above, it can be shown that the distribution of ξ itself is improper, the whole mass being concentrated at the origin, $\xi = 0$. But both these cases contradict the conditions of the theorem. Now the only alternative left is when both $c - a\beta$ and $\beta - \beta_0$ vanish simultaneously. But in this case we have $c = a\beta_0$, which is again contrary to the conditions of the theorem.

Now it may be noted that both $\varphi_1(u)$ and $\varphi(u, 0)$ may have real roots. Let ϵ and δ denote the smallest of the absolute values of the real roots of $\varphi_1(u)$ and $\varphi(u, 0)$, respectively. Since both $\varphi_1(u)$ and $\varphi(u, 0)$ are continuous functions of u and since $\varphi_1(0) = \varphi(0, 0) = 1$, it follows that $\epsilon > 0$ and $\delta > 0$. Then, restricting the values of a to an interval I, (a_1, a_2), for which $|a| < \epsilon/\delta$, we can always take the neighbourhood of the origin to be defined by $|u| < \delta$. Thus we have proved the existence of a neighbourhood $|u| < \delta$ of the origin and of an interval I, (a_1, a_2), such that both $\varphi_1(u)$ and $\varphi(u, 0)$ do not vanish if $a \varepsilon I$ and $|u| < \delta$.

Then, confining the values of u and a in these intervals, since the product $\varphi_1(au)\varphi(u, 0) \neq 0$, we may divide both sides of (8) by $\varphi_1(au)\varphi(u, 0)$ and thereby obtain

$$(c - a\beta)\frac{\varphi_1'(au)}{\varphi_1(au)} = (\beta - \beta_0)\frac{\varphi'(u, 0)}{\varphi(u, 0)}. \tag{12}$$

Next, integrating (12) with respect to u, we get

$$\ln \varphi_1(au) = \frac{a(\beta - \beta_0)}{c - a\beta} \ln \varphi(u, 0), \tag{13}$$

where the constant of integration vanishes by virtue of the fact that $\ln \varphi_1(0) = \ln \varphi(0, 0) = 0$.

Since the first moment of x exists, $\ln \varphi_1(au)$ is differentiable with respect to a in the interval (a_1, a_2). Thus it follows that $\theta(a) = a(\beta - \beta_0)/(c - a\beta)$ must also be differentiable with respect to a in the same interval; denoting this derivative by $\theta'(a)$, we may write

$$u \frac{\varphi_1'(au)}{\varphi_1(au)} = \theta'(a) \ln \varphi(u, 0). \tag{14}$$

Again, from the conditions of the theorem, $\theta(a) \neq 0$ for all a in the interval (a_1, a_2). Hence, using (12) and (14) together, we get

$$\begin{aligned} u \frac{\varphi'(u, 0)}{\varphi(u, 0)} &= a \frac{\theta'(a)}{\theta(a)} \ln \varphi(u, 0) \\ &= \lambda \ln \varphi(u, 0), \end{aligned} \tag{15}$$

where $\lambda = a\theta'(a)/\theta(a)$ for all a contained in the interval (a_1, a_2), and thus it follows evidently from (15) that λ is independent of a.

Then, excluding the origin from the interval $|u| < \delta$, that is, in the intervals $(0 < u < +\delta)$ and $(-\delta < u < 0)$, we may divide both sides of (15) by

$$u \ln \varphi(u, 0),$$

and obtain

$$\frac{1}{\ln \varphi(u, 0)} \cdot \frac{\varphi'(u, 0)}{\varphi(u, 0)} = \lambda \frac{1}{u}. \tag{16}$$

Hence, integrating (16) with respect to u, we get

$$\ln \ln \varphi(u, 0) = \begin{cases} \lambda \log |u| + \log c_1, & \text{for } 0 < u < +\delta \\ \lambda \log |u| + \log c_2, & \text{for } -\delta < u < 0. \end{cases} \tag{17}$$

Now, (17) evidently leads to the relation

$$\varphi(u, 0) = \begin{cases} e^{c_1 |u|^\lambda}, & \text{for } 0 < u < +\delta \\ e^{c_2 |u|^\lambda}, & \text{for } -\delta < u < 0, \end{cases} \tag{18}$$

where c_1 and c_2 are the constants of integration. But it is well known that necessary conditions for a function $\varphi(t)$ to be a characteristic function are:

(i) $\varphi(0) = 1$, (ii) $|\varphi(t)| \leq 1$, and (iii) $\varphi(-t) = \overline{\varphi(t)}$.

Hence, it evidently follows that c_1 and c_2 in (18) should be complex conjugates; that is, we may write

$$c_1 = -(A + iB) \quad \text{and} \quad c_2 = -(A - iB),$$

where $A \geqq 0$. Thus the formula,

$$\varphi(u, 0) = \exp\left[-\left(A + iB\frac{u}{|u|}\right)|u|^\lambda\right]$$

holds for all u in the interval $|u| < \delta$.

It can be easily shown that $\delta = +\infty$, since from the continuity of the characteristic function, we have $\varphi(\pm\delta, 0) \neq 0$, which contradicts the assumption that δ is the smallest of the absolute value of the real root of $\varphi(u, 0)$. Hence, the characteristic function of the distribution of ξ is given by

$$\varphi(u, 0) = \exp\left[-\left(A + iB\frac{u}{|u|}\right)|u|^\lambda\right]. \tag{19}$$

Now, it should be noted that the case $A = 0$ should be excluded, since when $A = 0$, $|\varphi(u, 0)| = 1$ for all u; this leads to the trivial case that the whole mass of the distribution is concentrated at a single point.

It is already pointed out in (15) that λ does not involve a, so that on solving $\lambda = a[\theta'(a)/\theta(a)]$ as a differential equation in a, we get

$$\theta(a) = K|a|^\lambda. \tag{20}$$

Hence, we have from (13)

$$\varphi_1(au) = \exp\left[-K\left(A + iB\frac{u}{|u|}\right)|au|^\lambda\right], \tag{21}$$

where $K > 0$ for the same reason as $A > 0$.

Next, we shall show that $1 < \lambda \leqq 2$. If $\lambda \leqq 1$, the first derivatives of both $\varphi_1(u)$ and $\varphi(u, 0)$ fail to exist at the origin, which means that the first moments of ξ and x do not exist, contrary to the assumption of our theorem. On the other hand, if $\lambda > 2$, the second derivatives of both the functions $\varphi_1(u)$ and $\varphi(u, 0)$ exist and vanish at the origin. In this case, the second moments of both ξ and x exist and are equal to zero, which means that the whole mass of the distribution is concentrated at the point $\xi = x = 0$, and we have $\varphi_1(u) = \varphi(u, 0) = 1$ for all u (c.f. Cramer [1]). Now, from Lévy and Khintchine [6], it follows evidently that the characteristic functions (19) and (21) uniquely determine the distribution function of a stable law with finite expectation when and only when the parameters A, B, K, and λ satisfy the restrictions

$$\begin{cases} A > 0; \quad K > 0; \quad 1 < \lambda \leqq 2; \\ \left|B \cos\left(\frac{\pi}{2}\lambda\right)\right| \leqq A \sin\left(\frac{\pi}{2}\lambda\right). \end{cases} \tag{22}$$

Proof of Sufficiency. We have to show that if the distribution functions of ξ and x are characterized by (19) and (21), respectively, with the parameters satisfying the restrictions as listed in (22), and further that if the first moment of η exists and the regression of η on ξ exists and is given by $E_\xi(\eta) = \beta_0\xi$, then the regression of $Y = cx + \eta$ on $X = ax + \xi$ should also be linear, where x is independent of the joint distribution of ξ and η.

Here we have

$$\ln \varphi_1 (au) = K|a|^\lambda \ln \varphi(u, 0). \tag{23}$$

Then we have

$$\frac{\varphi_1'(au)}{\varphi_1(au)} = \frac{1}{a} \cdot \frac{d \ln \varphi_1(au)}{du} = \frac{1}{a} K \mid a \mid^\lambda \frac{d \ln \varphi(u, 0)}{du},$$

so that we have

$$\varphi_1' (au) \varphi(u, 0) = \frac{1}{a} K|a|^\lambda \varphi_1(au) \varphi'(u, 0). \tag{24}$$

Then, if $\Phi(u, v)$ stands for the characteristic function of the joint cumulative distribution of X and Y, we have, on substituting the value of $\varphi_1'(au)\varphi(u, 0)$ as in (24) in (4) and (5) above,

$$\begin{cases} \left.\dfrac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = \left(\dfrac{c}{a} K |a|^\lambda + \beta_0\right) \varphi_1(au)\varphi'(u, 0), \\ \dfrac{d\Phi(u, 0)}{du} = (1 + K |a|^\lambda)\varphi_1(au)\varphi'(u, 0), \end{cases} \tag{25}$$

so that

$$\left.\frac{\partial\Phi(u, v)}{\partial v}\right]_{v=0} = \frac{\beta_0 + \dfrac{c}{a} K |a|^\lambda}{1 + K |a|^\lambda} \frac{d\Phi(u, 0)}{du}. \tag{26}$$

Then the proof follows at once from (26), using the lemma.

It is also interesting to note in this connection that if we further assume that either ξ or x has a finite variance, that is, that the second derivative of either (19) or (21) exists at the origin, then λ should be equal to 2, and hence both x and ξ should be normally distributed.

Corollary 1. (*The problem of Ragnar Frisch.*) *In the problem of Ragnar Frisch, which has been solved independently by Rao* [8], [9] *and by Fix* [3], *it has been assumed that x, ξ, and η are mutually independent chance variables. Thus, it may be treated as a special case of Theorem 2, above, by putting $\beta_0 = 0$.*

Corollary 2. (*Generalisation of Darmois' Theorem.*) *If x and y are two independent chance variables with finite expectations such that the regression of $Y = cx + dy$ $(d \neq 0)$ on $X = ax + by$ $(b \neq 0)$ exists and is linear for all a contained in a closed interval (a_1, a_2), where either $a_1 < a_2 < 0$ or $0 < a_1 < a_2$, and for some*

c for which the relationship $bc \neq ad$ is satisfied for all a in the interval, then both x and y should belong to the class of stable law with finite expectation.

This may be treated as a special case of Theorem 2, above, by taking $\xi = by$ and $\eta = dy$. Finally, we shall construct a simple counter-example to show that the theorem is not true when the regression of Y on X is linear for some fixed a.

For this purpose, let us take

$$\ln \varphi(t) = \int_0^\infty (\cos |t|\, x - 1) \frac{e^{-\sin^2 \ln x}}{x^{1+\delta}}\, dx \qquad (1 < \delta < 2). \tag{27}$$

We shall show that $\varphi(t)$ in (27) represents the characteristic function of a symmetric infinitely divisible law with a finite first moment which is assumed to be zero. First of all, we note easily that $\varphi(t)$ in (27), being real, represents the characteristic function of a symmetric law. Now, following the notations given by Loève [7], we define

$$G(x) = \begin{cases} -\displaystyle\int_x^\infty \frac{e^{-\sin^2 \ln x}}{x^{1+\delta}}\, dx & \text{if } x > 0, \\ \displaystyle\int_{-\infty}^x \frac{e^{-\sin^2 \ln |x|}}{|\, x\,|^{1+\delta}}\, dx & \text{if } x < 0, \end{cases}$$

where $1 < \delta < 2$.

It can be easily verified that $G(x)$ satisfies all the conditions stated in Loève's representation formula for the infinitely divisible law. Hence, $\varphi(t)$ in (27), above, is the characteristic function of a symmetric infinitely divisible law.

Using the transformation $|t|x = u$, (27) reduces to

$$\ln \varphi(t) = |t|^\delta \int_0^\infty (\cos u - 1) \frac{\exp\{-\sin^2[\ln u - \ln |t|\,]\}}{u^{1+\delta}}\, du, \qquad 1 < \delta < 2. \tag{28}$$

Now the first derivative of $\varphi(t)$ in (28) exists at the origin, so that the first moment exists and is equal to zero.

Again, from (28), we have

$$\ln \varphi(at) = |at|^\delta \int_0^\infty (\cos u - 1) \frac{\exp\{-\sin^2[\ln u - \ln |t| - \ln |a|\,]\}}{u^{1+\delta}}\, du, \qquad 1 < \delta < 2. \tag{29}$$

Then, using (28) and (29) together, we have at the point $a = e^{2k\pi}$, where k takes any of the values $\pm 1, \pm 2, \pm 3 \cdots$,

$$\ln \varphi(at) = |a|^\delta \ln \varphi(t). \tag{30}$$

If the characteristic functions of the cumulative distributions of x and ξ are given by (27), above, then proceeding exactly in the same way as in (23), (24), (25), and (26), it can be shown that the regression of $Y = cx + \eta$ on $x = ax + \xi$

is linear for some fixed a, where $a = e^{2k\pi}$ and k is any one of the numbers ± 1, ± 2, ± 3, $\cdots$.

In conclusion, the author expresses his thanks to the Referee for some helpful comments.

REFERENCES

1. H. Cramér, *Mathematical methods of statistics*, Princeton University Press, 1946, p. 91.
2. G. Darmois, "Sur une propriete caracteristique de la loi de probabilite de Laplace," *C. R. Acad. Sci. Paris*, Vol. 232 (1951), pp. 1999–2000.
3. E. Fix, "Distributions which lead to linear regressions," *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 79–91.
4. B. V. Gnedenko, "On a theorem of S. N. Bernstein," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 12 (1948), pp. 97–100.
5. M. Kac, "A characterization of the normal distribution," *Amer. J. Math.*, Vol. 61 (1939), pp. 726–728.
6. P. Lévy and A. Khintchine, "Sur les loi stables," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 374–376.
7. M. Loève, "Fundamental limit theorems of probability theory," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 329.
8. C. R. Rao, "Note on a problem of Ragnar Frisch," *Econometrica*, Vol. 15 (1947), pp. 245–249.
9. C. R. Rao, "A correction to note on a problem of Ragnar Frisch," *Econometrica*, Vol. 17 (1949), p. 212.
10. Colette Rothschild and Edith Mourier, "Sur les lois de probabilité à regression lineaire et écart type lié constant," Comptes Rendus 225 (1947), pp. 245–249.

NOTES

A VARIABLE PROBABILITY DISTRIBUTION FUNCTION[1]

By Ruric E. Wheeler

Howard College, Birmingham

1. Introduction and Summary. It is the purpose of this paper to develop an expression for the probability of x successes in n trials, $P(n, x)$, where the probability of success on a single trial depends both on the number of the trial and on the number of previous successes. This result should prove useful in obtaining various probability functions. It will be noted that this work includes the case considered by Woodbury [3].

2. Definitions. Letting $p_{r,s}$ be the probability of a success and $q_{r,s}$ the probability of a failure on the rth trial after s successes, with $p_{r,s} + q_{r,s} = 1$, we formulate the following definition.

Definition 1. We will use the symbol S_i to be a function of $p_{r,s}$ and x, where x is the number of successes, with the following defined properties:

(a) $S_i = \prod_{t=0}^{i-1} p_{t+1,t} \prod_{t=i}^{x-1} p_{t+2,t} q_{i+1,i}$, where $i \leq x$ ($\prod_i^{i-1}$ is defined to be 1).

(b) The product of S_i, S_j, and S_k in any order (or any number of factors) is defined to be

$$\prod_{t=0}^{i-1} p_{t+1,t} \prod_{t=i}^{j-1} p_{t+2,t} \prod_{t=j}^{k-1} p_{t+3,t} \prod_{t=k}^{x-1} p_{t+4,t} q_{i+1,i} q_{j+2,j} q_{k+3,k},$$

for $i \leq j \leq k \leq x$, where if S_t is a function of x_t successes ($t = i, j, k$), then the quantity x which appears in the formula for the product is the maximum of x_i, x_j and x_k. It should be noted that the product of S_i and S_j is not equal to the value of S_i multiplied by the value of S_j but is given by the above definition.

(c) $S_i^0 = \prod_{t=0}^{i-1} p_{t+1,t}$.

(d) We define $S_i(S_j + S_k)$ to be $S_iS_j + S_iS_k$, where the $(+)$ sign represents ordinary addition.

(e) S_i^r will represent the product $S_iS_iS_i \cdots S_i$ to r factors, and from (b), must be equal $\prod_{t=0}^{i-1} p_{t+1,t} \cdot \prod_{t=i}^{x-1} p_{t+r+1,t} \cdot \prod_{t=1}^{r} q_{t+i,i}$.

(f) Then from (a) and (e), $S_i^m S_j^n$ will be

$$\prod_{t=0}^{i-1} p_{t+1,t} \prod_{t=i}^{j-1} p_{m+t+1,t} \prod_{t=j}^{x-1} p_{m+n+t+1,t} \prod_{t=1}^{m} q_{i+t,i} \prod_{t=1}^{n} q_{m+j+t,j}.$$

Received February 23, 1953; revised October 5, 1955.

[1] These results were included in a dissertation submitted to the University of Kentucky in partial fulfillment of the requirements for the Ph.D. degree, June, 1952.

To illustrate this definition, we consider $S_2^3 S_3^1$, which is

$$\prod_{t=0}^{1} p_{t+1,t} \prod_{t=2}^{2} p_{3+t+1,t} \prod_{t=3}^{x-1} p_{4+t+1,t} \prod_{t=1}^{3} q_{2+t,2} \prod_{t=1}^{1} q_{3+3+t,3}$$

or

$$p_{1,0}\, p_{2,1}\, p_{6,2}\, q_{3,2}\, q_{4,2}\, q_{5,2}\, q_{7,3} \prod_{t=3}^{x-1} p_{t+5,t}\,.$$

We note that the multiplication of these symbols as defined follows the laws of positive integral and zero exponents.

Lemma 1. *The probability $P(n, x)$ can be expressed as a sum of products of S's for all n and x, such that $n \geqq x$.*

Proof. Let us consider x successes and $n - x$ failures in the following specified order. Suppose we have α_0 failures, then a success; α_1 failures, then a second success; α_2 failures, then a third success; etc.; finally, the xth success and then α_x failures, where $\alpha_0 + \alpha_1 + \alpha_2 + \cdots + \alpha_x$ must equal $n - x$. By theorems from elementary probability theory, the probability of x successes and $n - x$ failures in this specified order is

$$\prod_{t=1}^{\alpha_0} q_{t,0}\, p_{\alpha_0+1,0} \prod_{t=1}^{\alpha_1} q_{\alpha_0+t+1,1}\, p_{\alpha_0+\alpha_1+2,1} \cdots \prod_{t=1}^{\alpha_x-1} q_{\alpha_0+\alpha_1+\cdots+\alpha_{x-2}+t+x-1,x-1}$$

$$\cdot\, p_{\alpha_0+\cdots+\alpha_{x-1}+x,x-1} \prod_{t=1}^{\alpha_x} q_{\alpha_0+\cdots+\alpha_{x-1}+t+x,x}\,,$$

which we may write in terms of our defined symbols as $S_0^{\alpha_0} S_1^{\alpha_1} S_2^{\alpha_2} \cdots S_x^{\alpha_x}$. Since $P(n, x)$ is the sum of terms such as this, it can be expressed as a sum of products of S's for all n and x such that $n \geqq x$.

3. Development of $P(n, x)$. Let us consider the following partial difference equation

$$P(n, x) = p_{1,0}P''(n - 1, x - 1) + q_{1,0}P'(n - 1, x), \tag{1}$$

where $P''(n - 1, x - 1)$ represents the probability of $x - 1$ successes in $n - 1$ trials with the probability of success on the first of the $n - 1$ trials being $p_{2,1}$, and where $P'(n - 1, x)$ is the probability of x successes in $n - 1$ trials with the probability of success on the first of the $n - 1$ trials being $p_{2,0}$. The boundary conditions for this equation are $P(n, x) = 0$ for $x < 0, x > n$, and $P(0, 0) = 1$. Using the generating function $G(x, \theta) = \sum_{k=x}^{\infty} P(k, x)\theta^k$, one may obtain, under the given boundary conditions, a difference equation involving generating functions. From (1) we have that

$$\sum_{k=x}^{\infty} P(k, x)\theta^k = \sum_{k=x}^{\infty} p_{1,0}P''(k - 1, x - 1)\theta^k + \sum_{k=x}^{\infty} q_{1,0}P'(k - 1, x)\theta^k,$$

which in turn gives

$$G(x, \theta) = p_{1,0}\theta G''(x - 1, \theta) + q_{1,0}\theta G(x, \theta). \tag{2}$$

Now let us make use of the displacement operator E. Considering $p_{i,j}$ as a function of i and j, we will use E operating on $p_{i,j}$ to be $p_{i+1,j+1}$. From the properties of E as given in [2],

$$E(S_i) = E\left(\prod_{t=0}^{i-1} p_{t+1,t} \prod_{t=i}^{x-1} p_{t+2,t}\, q_{i+1,i}\right),$$

or

$$E(S_i) = \prod_{t=1}^{i} p_{t+1,t} \prod_{t=i+1}^{x} p_{t+2,t}\, q_{i+2,i+1}\,.$$

Thus, $p_{1,0}E(S_i) = \prod_{t=0}^{i} p_{t+1,t} \prod_{t=i+1}^{x} p_{t+2,t}\, q_{i+2,i+1} = S_{i+1}$, where if S_i is a function of x successes, then S_{i+1} is a function of $x + 1$ successes. Likewise,

$$p_{1,0}E(S_iS_j \cdots S_v) = S_{i+1}S_{j+1} \cdots S_{v+1}$$

and $p_{1,0}E(a_iS_i + a_jS_j + \cdots) = a_iS_{i+1} + a_jS_{j+1} + \cdots$, where the a's are constants.

We will now simplify difference equation (2) by showing that $G''(x, \theta) = EG(x, \theta)$ and $q_{1,0}G'(x, \theta) = S_0G(x, \theta)$. Since these generating functions are power series in θ, it will be sufficient to show that coefficients of like powers of θ in each equation are equal. $P''(n, x)$ has exactly the same form as $P(n, x)$, but has the subscripts of each p and q increased by one (i.e., each $p_{i,j}$ in $P(n, x)$ is $p_{i+1,j+1}$ in $P(n, x)$). From the definition and the properties of E, it is evident that $EP(n, x) = P''(n, x)$. Thus, since $P(n, x)$ and $P''(n, x)$ are the coefficients of θ^n in $G(x, \theta)$ and $G''(x, \theta)$, respectively, we have $EG(x, \theta) = G''(x, \theta)$.

In a like manner, from the properties of S and the definition of $P'(n, x)$, it follows that S_0 multiplied by $P(n, x)$ equals $q_{1,0}P'(n, x)$ and hence $S_0G(x, \theta) = q_{1,0}G'(x, \theta)$.

By making these substitutions, difference equation (2) becomes

$$G(x, \theta) - S_0G(x, \theta)\theta = p_{1,0}EG(x - 1, \theta)\theta$$

or $(1 - S_0\theta)G(x, \theta) = EG(x - 1, \theta)\theta$. Multiplying both sides of this equation by $1 + S_0\theta + S_0^2\theta^2 + \cdots$, we obtain

$$G(x, \theta) = (1 + S_0\theta + S_0^2\theta^2 \cdots)EG(x - 1, \theta)\theta.$$

Then, using $1/(1 - S_0\theta)$ to represent $1 + S_0\theta + S_0^2\theta^2 + \cdots$, we have

$$G(x, \theta) = \frac{\theta}{1 - S_0\theta} EG(x - 1, \theta). \tag{3}$$

Since $G(0, \theta) = 1/(1 - S_0\theta)$, the solution of (3) becomes

$$G(x, \theta) = \frac{\theta^x}{(1 - S_0\theta)(1 - S_1\theta) \cdots (1 - S_x\theta)}. \tag{4}$$

For ordinary multiplication, it is shown on page 313 of [3] that the coefficient of θ^n in an expansion similar to this one is the xth divided difference of S_0^n, a

polynomial of degree $n - x$ in the S's. Since our function S satisfies the laws of multiplication for positive integral and zero exponents, the coefficient of θ^n can thus be expressed as the xth divided difference of S_0^n. Or the probability of x successes in n trials is $P(n, x) = \Delta^x S_0^n$, where the symbol Δ is the divided difference.

It is of interest to see what happens to this expression when the probability of success on a single trial depends only on the number of previous successes. Since $P(n, x) = \Delta^x S_0^n$, any term of this polynomial may be written as $S_0^{n_0} S_1^{n_1} \cdots S_x^{n_x}$, where the n_i may take on values $0, 1, 2, \cdots, n - x$. Since our $p_{i,j}$ is now restricted so that it depends only on the number of previous successes (omitting the first subscripts), each term becomes $p_0 p_1 \cdots p_{x-1} q_0^{n_0} q_1^{n_1} \cdots q_x^{n_x}$, or $P(n, x) = p_0 p_1 \cdots p_{x-1} \Delta^x q_0^n$. This is the expression for $P(n, x)$ that was obtained by Woodbury in [3].

By specifying the exact law by which the probability of success on a single trial changes from trial to trial, we may obtain probabilities that determine various desired distributions. As an example, let the probability of success on the first trial be expressed as $p/1$, and let the numerator of this fraction be increased by λ for each success and the denominator increased by λ for each trial. For this distribution, we obtain, from Definition 1,

$$S_0^r = \frac{p}{1 + r\lambda} \frac{p + \lambda}{1 + (r + 1)\lambda} \cdots \frac{p + (x - 1)\lambda}{1 + (x + r - 1)\lambda} \frac{q}{1} \frac{q + \lambda}{1 + \lambda} \cdots \frac{q + (r - 1)\lambda}{1 + (r - 1)\lambda}$$

and

$$S_i^r = \frac{p}{1} \frac{p + \lambda}{1 + \lambda} \cdots \frac{p + (i - 1)\lambda}{1 + (i - 1)\lambda} \frac{p + i\lambda}{1 + (r + i)\lambda} \cdots$$
$$\frac{p + (x - 1)\lambda}{1 + (r + x - 1)\lambda} \frac{q}{1 + i\lambda} \frac{q + \lambda}{1 + (i + 1)\lambda} \cdots \frac{q + (r - 1)\lambda}{1 + (i + r - 1)\lambda}.$$

Since S_0^r is equal to S_i^r for $i = 1, 2, 3, \cdots, x - 1$, and since $\Delta^x S_0^n$ has C_x^n terms, then $P(n, x) = C_x^n S_0^{n-x}$, or

$$C_x^n \frac{p(p + \lambda) \cdots (p + [x - 1]\lambda)(q)(q + \lambda) \cdots (q + [n - x - 1]\lambda)}{(1 + [n - 1]\lambda)!},$$

which is the probability of exactly x successes in n trials for the Polya distribution as given in [1].

REFERENCES

[1] W. Feller, *Probability Theory and Its Applications*, John Wiley and Sons, New York, 1950, p. 128.

[2] C. Jordan, *Calculus of Finite Differences*, Chelsea Publishing Co., New York, 1947, pp. 5 and 18.

[3] Max A. Woodbury, "On a Probability Distribution", *Ann. Math. Stat.* Vol. 20 (1949), pp. 311–315.

UNIFORM CONVERGENCE OF RANDOM FUNCTIONS WITH APPLICATIONS TO STATISTICS[1]

By Herman Rubin

Stanford University[2]

0. Introduction and Summary. In many statistical problems, we obtain functions of both the random variable and the parameters involved, from whose asymptotic behavior we may deduce the asymptotic behavior of certain estimates. In many of these cases, it is sufficient to demonstrate uniform convergence with probability one of these functions. In this paper, a set of sufficient conditions for this is given, and we show how these results may be applied to some statistical problems.

1. Statement of the theorem.[3] Let $X_1, \cdots, X_n, \cdots$ be a sequence of independent and identically distributed variables with values in an arbitrary space X. Let T be a compact topological space, and let f be a complex-valued function on $T \times X$, measurable in x for each $t \in T$. Let P be the common distribution of the X_i.

Theorem 1. *If there is an integrable g such that $|f(t, x)| < g(x)$ for all $t \in T$ and $x \in X$, and if there is a sequence S_i of measurable sets such that*

$$P(X - \textstyle\bigcup_{i=1}^{\infty} S_i) = 0,$$

and for each i, $f(t, x)$ is equicontinuous in t for $x \in S_i$, then with probability one,

$$\frac{1}{n}\sum_{k=1}^{n} f(t, X_k) \to \int f(t, x)\, dP(x)$$

uniformly jor $t \in T$, and the limit function is continuous.

We may assume the S_i are monotonically increasing. Let $\epsilon > 0$ be given. Then for some i, $\int_{X-S_i} g(x)\, dP(x) < \epsilon/5$.

Since $f(t, x)$ is equicontinuous in t for $x \in S_i$ and T is compact, there exist $t_1, \cdots, t_q$ and open subsets $N_1, \cdots, N_q$ of T such that $\bigcup_{j=1}^{q} N_j = T$, $t_j \in N_j$, and for $t \in N_j$ and $x \in S_i$, $|f(t, x) - f(t_j, x)| < \epsilon/4$. Let $Y_{jk} = f(t_j, X_k)$; $Z_k = g(X_k)$, $X_k \not\in S_i$; $Z_k = 0$, $X_k \in S_i$.

By the strong law of large numbers, we may select an N such that, if $A_j = \mathcal{E}(Y_{jk})$ and $\delta > 0$,

$$P\left(\text{for some } n > N, \quad \left|\frac{1}{n}\sum_{k=1}^{n} Y_{jk} - A_j\right| \geqq \epsilon/4\right) < \delta/2q, \qquad j = 1, \cdots, q,$$

Received November 8, 1952.

[1] Research performed under an ONR contract. Presented to the Institute of Mathematical Statistics, June 15, 1951.

[2] Present address, University of Oregon.

[3] An essentially equivalent theorem was proved by Le Cam in [2].

and

$$P\left(\text{for some } n > N, \quad \left|\frac{1}{n}\sum_{k=1}^{n} Z_k\right| \geq \epsilon/4\right) < \delta/2.$$

But if $t \varepsilon N_j$, $|f(t, X_k) - Y_{jk}| < \epsilon/4 + 2Z_k$. Hence,

$$P\left(\text{for some } n > N \text{ and some } t, t \varepsilon N_j \text{ and } \left|\frac{1}{n}\sum_{k=1}^{n} f(t, X_k) - A_j\right| \geq \epsilon\right) < \delta.$$

Therefore, $1/n \sum_{k=1}^{n} f(t, X_k)$ converges uniformly to a continuous function with probability one. By the strong law of large numbers, that function is

$$\int f(t, x)\, dP(x).$$

2. Applications. As an application of this theorem, we see that the sample characteristic function converges to the population characteristic function uniformly with probability one in any bounded interval, since $f(t, x) = e^{itx}$ satisfies the conditions of the theorem.

It may happen that log $L(x \mid \theta) = f(x, \theta)$ satisfies the conditions of the theorem. For example, for the multivariate normal, the Poisson, Cauchy, χ^2, double exponential, and many other distributions, we are led to the almost certain convergence of maximum likelihood estimates to the true values if the parameter is restricted to a compact set.

More difficult estimation procedures can also be shown to be consistent. For example, consider a problem of Reiersøl [4]. The model is

$$x_i = \xi_i \cos \alpha + u_i,$$

$$y_i = \xi_i \sin \alpha + v_i,$$

where u_i and v_i have a joint normal distribution, ξ_i is not normal, and ξ_i, (u_i, v_i) are independent. Let $\rho = t \sin \beta$, $\sigma = -t \cos \beta$, and let

$$\varphi(t, \beta, X_j) = e^{i\rho x_j + i\sigma y_j}.$$

Then $1/n \sum_{j=1}^{n} \varphi(t, \beta, X_j) \to \psi(t, \beta)$ uniformly with probability one for t in any finite interval. Let

$$\chi(\beta) = \int_{-\infty}^{\infty} |\psi(2t, \beta) - \psi^3(t, \beta)\psi(-t, \beta)|^2\, d\lambda(t),$$

where λ is a bounded monotone function, such that for any $\epsilon > 0$, $\lambda(\epsilon) - \lambda(-\epsilon) > 0$. Then χ is a periodic function of period π, and $\chi(\beta) = 0$ only for $\beta = \alpha + k\pi$. Let

$$\psi_n(t, \beta) = \frac{1}{n}\sum_{j=1}^{n} \varphi(t, \beta, X_j),$$

$$\chi_n(\beta) = \int_{-\infty}^{\infty} |\psi_n(2t, \beta) - \psi_n^3(t, \beta)\psi_n(-t, \beta)|^2\, d\lambda(t).$$

Since ψ_n is bounded, it follows that $\chi_n(\beta) \to \chi(\beta)$ uniformly with probability one. Hence, if b_n minimizes $\chi_n(\beta)$, it follows that $b_n \to \beta$ with probability one, in the sense of convergence mod π.

This result is stronger than that obtained by Neyman about Reiersøl's problem. The method can also be extended to Neyman's extension of the problem [3].

We can, in fact, obtain some very strong results on the existence of consistent estimates.

THEOREM 2. *Let $\mathfrak{F}$ be a family of distributions on Euclidean n-space. Let π be a continuous function mapping $\mathfrak{F}$ into the topological space $\mathcal{P}$. Then there exists a sequence p_k of functions on Euclidean kn-space to $\mathcal{P}$ such that if $X_1, \cdots, X_k, \cdots$ are independently distributed with distribution function $F \varepsilon \mathfrak{F}$ then $\lim_{k\to\infty} p_k(X_1, \cdots, X_k) = \pi(F)$ with probability one.*

In other words, any continuous parameter is consistently estimable. The converse is not true, since moments, which are not continuous parameters, are consistently estimable. It is an unsolved problem, which functions π of a distribution in a family $\mathfrak{F}$ are consistently estimable—even whether the topological structure of the family $\mathfrak{F}$ and the topological properties of the function π are sufficient to characterize consistently estimable parameters π.

Let us proceed to the proof of the theorem.

Let λ be a non-negative finite measure on Euclidean n-space such that every open set has positive measure. For each $F \varepsilon \mathfrak{F}$, let

$$\psi(t_1, \cdots, t_n, F) = \mathcal{E}\left(\exp i \sum_{j=1}^{n} t_j x_j \mid F\right),$$

i.e., $\psi(t_1, \cdots, t_n, F)$ is the characteristic function of F evaluated at $t_1, \cdots, t_n$.

Similarly, let

$$\psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) = \frac{1}{k}\sum_{h=1}^{k} \exp\left(i \sum_{j=1}^{n} t_j X_{hj}\right).$$

Then define

$$\rho_k^2(X_1, \cdots, X_k)$$
$$= \inf_{F \varepsilon \mathfrak{F}} \int |\psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F)|^2 \cdot d\lambda(t_1, \cdots, t_n)$$

and let p_k be any function such that for every $X_1, \cdots, X_k$ there is an $F_k \varepsilon \mathfrak{F}$ satisfying

$$\int |\psi_k(t_1, \cdots t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F_k)|^2 \, d\lambda(t_1, \cdots, t_n)$$
$$< \rho_k^2(X_1, \cdots, X_k) + \frac{1}{k}$$

and

$$p_k(X_1, \cdots, X_k) = \pi(F_k).$$

From Theorem 1, we see that $\psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k)$ approaches $\psi(t_1, \cdots, t_n, F)$ uniformly with probability one for $t_1, \cdots, t_n$ in any bounded set. Therefore,

$$\int | \psi_k(t_1, \cdots, t_n, X_1, \cdots, X_k) - \psi(t_1, \cdots, t_n, F) |^2 d\lambda(t_1, \cdots, t_n)$$

approaches zero with probability one. Hence,

$$\int | \psi(t_1, \cdots, t_n, F_k) - \psi(t_1, \cdots, t_n, F) |^2 d\lambda(t_1, \cdots, t_n)$$

approaches zero with probability one, and thus $F_k \rightarrow F$ almost certainly. The result then follows from the continuity of π.

Similar results can be obtained in the case of a continuously identified parameter. If a structure S generates the distribution F, we may ask whether a function φ defined on the space $\mathcal{S}$ of structures is determined by the distribution $\mathfrak{F}$. If so, we say [1] that φ is *identified* at F.

Let us formulate the preceding definition without regard to the structure S. We obtain for each F in a class $\mathfrak{F}$ of distributions, a non-null set $\Phi(F)$ in the parameter space. The condition that Φ is identified at F then becomes that $\Phi(F)$ has one element.

Let us say that Φ is *continuously identified* at F with respect to $\mathfrak{F}$ if for every sequence $F_1, \cdots, F_n, \cdots$ of distributions of $\mathfrak{F}$ such that $F_n \rightarrow F$, and for any sequence $\theta_1, \cdots, \theta_n, \cdots$ such that $\theta_k \varepsilon \Phi(F_k)$ for all k, θ_n converges to the one element of $\Phi(F)$.

Then by a method similar to that of Theorem 2 we obtain

THEOREM 3. *Let $\mathfrak{F}$ be a family of distributions on Euclidean n-space and let Φ map $\mathfrak{F}$ into the set of non-null subsets of $\mathcal{P}$. Then there exists a sequence p_k of functions on Euclidean kn-space to $\mathcal{P}$ such that for any $F \varepsilon \mathfrak{F}$, if $X_1, \cdots, X_k, \cdots$ are independently distributed with distribution function F and Φ is continuously identifiable at F with respect to $\mathfrak{F}$, then $p_k(X_1, \cdots, X_k)$ approaches the element of $\Phi(F)$ with probability one.*

REFERENCES

[1] L. HURWICZ, "Generalization of the concept of identification," *Statistical Inference in Dynamic Economic Models*, John Wiley and Sons, New York, 1950, pp. 245–257.

[2] L. LE CAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates", *University of California Publications in Statistics*, Vol. 1 (1953), pp. 277–300.

[3] J. NEYMAN, "Existence of consistent estimates of the directional parameter in a linear structural relation between two variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 497–512.

[4] O. REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–389.

CORRECTION TO "ON THE MAXIMUM NUMBER OF CONSTRAINTS OF AN ORTHOGONAL ARRAY"

By Esther Seiden

Northwestern University

The proof of Lemma 2 of the paper mentioned in the above title (*Ann. Math. Stat.* Vol. 26 (1955), pp. 132–135) is incorrect. The number 20 on top of page 134 should be replaced by 15 and hence no contradiction has been reached with $n_{12}^6 = 45$. Fortunately the assertion made in the above mentioned remains valid. The last seven lines of page 133 and the first two lines of page 134 should be deleted and replaced by the following:

This means that every 4-rowed orthogonal subarray must satisfy the equality $n_{14}^4 = 1$, contrary to Lemma 1 of the paper "Further remark on the maximum number of constraints of an orthogonal array" (to appear in the December issue, *Ann. Math. Stat.* Vol. 26 (1955), which asserts that no such array exists.

I wish to thank W. S. Connor for pointing out the mistake in my former proof.

ABSTRACTS OF PAPERS

(*Abstracts of papers presented at the New York meeting of the Institute, December 27–30, 1955*)

1. The Midrange of a Sample as an Estimator of the Population Midrange, Paul R. Rider, Wright-Patterson Air Force Base.

A study is made of the distribution of the midranges of samples from five different symmetric populations of limited range, and of the relative efficiency of midrange and mean in estimating the population midrange, or mean, or median. It is found that the midrange is more efficient than the mean for all of the populations considered, and that this efficiency increases as the standardized fourth moment decreases.

2. Distribution of the Product of Maximum Values in Samples from a Rectangular Population, Paul R. Rider, Wright-Patterson Air Force Base, (By Title).

The distribution of the product of maximum values in samples from a rectangular distribution is derived. Results are obtained for the case of two samples of different sizes and for k samples of the same size.

3. A Note on Non-Recurrent Random Walks, Cyrus Derman, Columbia University, (By Title).

Let $\{X_i\}$, $i = 1, \cdots$, be a sequence of independent and identically distributed random variables with density function $f(x)$ and $EX_i = \lambda > 0$. Let $\{S_n\}$, $n = 1, \cdots$, be the sequence of cumulative sums $S_n = \sum_{i=1}^{n} X_i$, $H(x) = \sum_{n=1}^{\infty} P(S_n < x)$, and $h(x) = H'(x)$. Let A be any Borel set of the positive real numbers and $m(A)$ denote its Lebesgue measure.

Chung and Derman (to appear in the *Pacific J. Math.*) proved that (I) $\Phi(A) = P(S_n \varepsilon A$ infinitely often$) = 0$ if $m(A) < \infty$, and (II) that $\Phi(A) = 1$ if $m(A) = \infty$, provided that as $x \to \infty$, $0 < \liminf h(x) \leqq \limsup h(x) < \infty$. The following theorem was proved: (i) If $\lambda < \infty$ and $\limsup h(x) < \infty$, then $\liminf h(x) > 0$ and consequently (I) and (II) hold. (ii) If $\liminf h(x) > 0$, then (I) holds. (iii) If $\lambda < \infty$, and if there exists a constant $\alpha > 0$ and an interval (a, b) such that $f(x) \geqq \alpha$ for $x \varepsilon (a, b)$, then $\liminf h(x) > 0$.

4. **Statistical Spectral Analysis, I: Consistent Asymptotically Normal Estimates of the Covariance Function and Spectral Averages,** EMANUEL PARZEN, Columbia University, (By Title).

Let the wide-sense stationary time series $x(t)$ have mean m, covariance $R(v) = E\, x(t)\, x(t+v) - m^2$, spectral distribution function $F(\omega)$ such that $R(v) = \int e^{iv\omega}\, dF(\omega)$, and spectral density function $f(\omega) = F'(\omega)$. The problem of statistical spectral analysis is to estimate these quantities on the basis of an observed sample. We shall be especially concerned with finding consistent and asymptotically normal estimators, for both continuous and discrete parameter processes, under the following *assumptions*: (1) $R(v)$ is absolutely and square summable; (2) the process $y(t) = x(t) - m$ is stationary of order 4; (3) the non-Gaussian part of the fourth moment, or the fourth cumulant, $Q(v_1, v_2, v_3) = E\, y(t)\, y(t+v_1)\, y(t+v_2)\, y(t+v_3) - R(v_1)\, R(v_2 - v_3) - R(v_2)\, R(v_3 - v_1) - R(v_3)\, R(v_1 - v_2)$ is absolutely summable; (4) there is an absolutely integrable function $g(\omega_1, \omega_2, \omega_3)$ such that $Q(v_1, v_2, v_3) = \iiint d\omega_1\, d\omega_2\, d\omega_3 \exp i\, [\omega_1 v_1 + \omega_2 v_2 + \omega_3 v_3]\, g(\omega_1, \omega_2, \omega_3)$. Examples are given of processes satisfying these assumptions; they are examples of multilinear processes. Given $x(t)$, for $0 \leqq t \leqq T$ (or for $t = 1, \cdots, T$ in the discrete case), define the sample mean m_T, the sample covariance $R_T(v)$, the sample spectral density (or periodogram) $f_T(\omega)$, and sample spectral averages $J_T(A)$ by $Tm_T = \int_0^T x(t)\, dt$, $TR_T(v) = \int_0^{T-|v|} [x(t) - m_T]\, [x(t + |v|) - m_T]$, $R_T(v) = \int e^{iv\omega} f_T(\omega)\, d\omega$, $J_T(A) = \int A(\omega)\, f_T(\omega)\, d\omega$ for suitable $A(\omega)$. Expressions are obtained for the limit, as $T \to \infty$, of $TE(m_T - m)^2$, $TE\, |R_T(v) - R(v)|^2$, and $E\, |J_T(A) - J(A)|^2$, where $J(A) = \int A(\omega)\, f(\omega)\, d\omega$.

5. **Statistical Spectral Analysis, II: Asymptotic Mean Square Error of a Class of Estimates of the Spectral Density,** EMANUEL PARZEN, Columbia University.

It is well known that the sample spectral density (or periodogram) $f_T(\omega)$ is not a consistent estimate of the spectral density. A class of consistent estimates may be found in the following way (where we write out the formulas only for the discrete parameter case, noting that similar statements hold for the continuous case). Define $f_T^*(\omega) = (\frac{1}{2\pi}) \times \sum_{|v| \leq T} e^{-iv\omega}\, k(B_T v)\, R_T(v)$, where B_T is a sequence of constants such that $B_T \to 0$ and $TB_T \to \infty$ as $T \to \infty$, and the kernel $k(u)$ is defined for all real u as an even, bounded, square integrable function, with Fourier transform $K(\omega)$. It is assumed to satisfy $k(0) = 1$, and $\int_{-T}^{T} |k(u)|\, du \leqq MT^{\frac{1}{2}-\epsilon}$ for some $\epsilon > 0$ and constant M. Various estimates that have been proposed for the spectral density (Bartlett, Daniell, Grenander, Tukey) may be regarded as special instances of $f_T^*(\omega)$. To study the properties of $f_T^*(\omega)$, and to form a theory of the optimum estimate of

this form, we need to know the mean square error $E \mid f_T^*(\omega) - f(\omega) \mid^2$. An asymptotic expression for it can be obtained from the following two theorems.

Theorem I: $TB_T\sigma^2[f_T^*(\omega)] \to \mid f(\omega) \mid^2 \int k^2(u)\, du\ \{1 + \delta(0, \omega)\}$ *where* $\delta(0, \omega) = 1$ or 0 according as $\omega = 0$ or $\neq 0$.

Theorem II: *Let* $r > 0$ *be such that* $\Sigma \mid v \mid^r \mid R(v) \mid < \infty$, *and* $k^{(r)} = \lim_{u \to 0} (1 - k(u))/ \mid u \mid^r$ *is finite. Then* $B_T^{-2r} \mid E f_T^*(\omega) - f(\omega) \mid^2 \to \mid k^{(r)} f^{(r)}(\omega) \mid^2$ *where* $f^{(r)}(\omega) = (\frac{1}{2}\pi)\Sigma e^{-iv\omega} \mid v \mid^r R(v)$.

6. A Central Limit Theorem for Multilinear Stochastic Processes, Emanuel Parzen. Columbia University, (By Title).

A definition of a multilinear process is given. Intuitively, a stochastic process $x(t)$, defined for all real t, is said to be multilinear if it arises from a process with independent increments by means of passage through a finite bank of "linear filters" and "polynomial law instantaneous devices". Many physically observed stochastic processes may be assumed to arise in this way. Let $S_T = \int_0^T x(t)\, dK(t)$ and $S_N' = \Sigma_n^N x(t_{n-1})\, a_n$, for some sequence of points $t_n \to \infty$, constants a_n, and weighting function $K(t)$. Conditions are given in terms of moments, in order that the normalized random variables $(S_T - ES_T)/\sigma S_T$ and $(S_N' - ES_N')/\sigma S_N'$ tend in distribution to a normal law with zero mean and unit variance.

7. An Extension of Cramér's Theorem 20.6 to Random Functions with Values in a Metric Space, Emanuel Parzen, Columbia University, (By Title).

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $(R, \rho; \mathcal{B})$ be a metric measurable space, by which we mean that R is metrized by ρ, and $\mathcal{B}$ is the minimal σ-field over the open sets in $\mathcal{B}$. A random function X on Ω to R is $\mathcal{A}$-measurable if $\mathcal{A}$ contains the inverse image under X of every set in $\mathcal{B}$; then X generates a probability P_X on $\mathcal{B}$. For a sequence of $\mathcal{A}$-measurable functions X_n, define X_n to converge in distribution to X (denoted $X_n \xrightarrow{D} X$) if, for every bounded real-valued $\mathcal{B}$-measurable function $f(x)$ on R whose set of discontinuities is of P_x-measure 0, (*) $\int f(X_n)\, dP \to \int f(X)\, dP$. By a result of P. P. Billingsley (Ph.D. thesis, Princeton, 1955; Th. 1.1), $X_n \xrightarrow{D} X$ if, and only if, (*) holds for every bounded uniformly continuous function on R. Note also that, if X is a constant, then $X_n \xrightarrow{D} X$ if, and only if, $\rho\,(X_n, X) \to 0$ in probability. Extension of Theorem 20.6 in Cramér (*Mathematical Methods of Statistics*, Princeton, 1946): Let X_n, X be $\mathcal{A}$-measurable functions to $(R_1, \rho_1; \mathcal{B}_1)$ and let Y_n, Y be $\mathcal{A}$-measurable function to $(R_2, \rho_2; \mathcal{B}_2)$. Let R be the Cartesian product of R_1 and R_2, and let ρ be a metric on R which agrees with the metrics ρ_1 and p_2, and such that (X_n, Y_n), (X, Y) are $\mathcal{A}$-measurable to $(R, \rho; \mathcal{B})$. Suppose $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{D} Y$, and Y is a constant. Then $(X_n, Y_n) \xrightarrow{D} (X, Y)$. *Proof.* It suffices to show $\int f(X_n, Y_n)\, dP \to \int f(X, Y)\, dP$ for any bounded uniformly continuous function $f(x, y)$ on R. Clearly, $\int f(X_n, Y)\, dP \to \int f(X, Y)\, dP$. Let $\delta(\epsilon)$ be such that $\mid f(X_n, Y_n) - f(X_n, Y) \mid < \epsilon$ for $\rho_2(Y_n, Y) < \delta(\epsilon)$. Since $\int \mid f(X_n, Y_n) - f(X_n, Y) \mid dP \leq \epsilon + P\{\rho_2(Y_n, Y) > \delta(\epsilon)\}$, the desired conclusion may now be inferred.

8. Orthogonality and Fractional Replication of Factorial Experiments, Allan Birnbaum, Columbia University.

A simple characterization of orthogonal factorial designs is derived from the condition for orthogonality of appropriate vector subspaces of the sample space. This leads naturally to: (a) the definitions of various classes of orthogonal designs, some of them standard (e.g., Latin squares) and some less familiar (e.g., "Latin rectangles"); (b) some lower bounds on the fraction of replication which is consistent with orthogonality; (c) some elementary methods of construction of orthogonal fractional replicates, which in some cases can be shown to consist of a smallest possible fractional. Examples of such fractional replicates, including cases of factors at unequal numbers of levels, are given.

9. On the Second Sample Size Function of a Bayes Two-Stage Test for the Mean, Morris Skibinsky, Purdue University.

This paper investigates in detail the second sample size function of a Bayes two-stage rule that decides between two possible values for the mean of a normal distribution which has unit variance. The second sample size is the greatest integer less than or equal to a number, $\hat{y}(\text{const.} \times \log (r_m/Wg), W, M)$, where $\hat{y}(t, \gamma, \mu)$ is, for fixed values of its arguments, a value of y for which a certain function, $U(y, t, \gamma, \mu)$, is absolutely minimum; m is the size of the first sample, r_m the value of the probability ratio from the first sample; W and g the ratios, respectively, of the simple wrong decision losses and the a priori probabilities associated with the two possible means; and M is the minimum wrong decision loss. Certain monotonicity, symmetry, and continuity properties of $\hat{y}$, and functions related to it, are proved, and an asymptotic expression for the function is found when the minimum wrong decision loss is large. A subsequent paper, continuing this investigation, will consider Bayes two-stage rules having optimum properties with respect to expected overall sample size among rules of the same power.

10. A New Estimation Procedure for a Linear Combination of Exponentials, (Preliminary Report), Richard G. Cornell, Oak Ridge National Laboratory and Virginia Polytechnic Institute.

A new estimation procedure is developed for the parameters of the model $y_{ij} = \alpha_1 e^{-\lambda_1 t_i} + \alpha_2 e^{-\lambda_2 t_i} + \cdots + \alpha_p e^{-\lambda_p t_i} + e_{ij}$. The errors e_{ij} are independently and normally distributed about mean zero with equal variances. The parameters λ_k are restricted to be positive and the observation points t_i are equally spaced. The number of points at which observations are taken is specified to be an integral multiple of the number of parameters. Also, equal numbers of observations are required at each observation point. Estimates are obtained by forming as many independent sums from the observations y_{ij} as there are parameters, equating these sums to their expectations, and solving for estimates of the parameters. A computationally simple, non-iterative solution is found. The resultant estimators are not only asymptotically normally distributed, but are also consistent, sufficient and asymptotically efficient. The limiting properties are demonstrated as either the number of observation points or the number of observations per observation point grows infinitely large.

11. A Note on Weighted Randomization, D. R. Cox, University of North Carolina.

The standard methods of randomization used in experimental design consist of selecting an arrangement at random from a set S of similar arrangements, giving each arrangement

in the set equal chance of selection. As is well known this device makes the standard designs unbiased in the sense that, under weak assumptions, the randomization expectations of linear and quadratic functions of the observations agree with their values as calculated from an appropriate linear model with random residuals. It is also known that these results do not hold when an adjustment for concomitant variation is made by analysis of covariance. In the present paper it is pointed out that a randomization justification for the covariance procedure can be provided if weighted randomization is used, i.e. if the arrangement for use is selected giving different arrangements in the set S appropriate unequal chances of selection. Possible applications are considered briefly.

12. On the Analysis of Incomplete Block Designs, MARVIN ZELEN, National Bureau of Standards.

Let there be 2v normal populations which can be divided into two sets of v populations each, such that the unknown parameters of each set are (μ_i, σ_1^2) and (μ_i, σ_2^2) $i = 1, 2, \cdots, v$. Consider the null hypothesis $H_0: (\mu_i = 0 \text{ for } i = 1, 2, \cdots, v)$ against the alternative hypothesis $H_1: (\mu_i \neq 0 \text{ for } i = 1, 2, \cdots, v)$. If a sample of size r_j is made for each of the v populations of the jth set $(j = 1, 2)$, then one can test H_0 using two independent F-ratios. The main problem is to combine the two independent tests of significance into one single test having (perhaps) greater power than either of the individual tests. This problem arises in the analysis of incomplete block designs where one set corresponds to the intra-block analysis and the other to the inter-block analysis. The object of this paper is to show that exact statistical tests do exist for combining intra- and inter-block information. Methods are discussed for combining the two tests and a comparison of the power function is made for particular numerical values of the alternative hypothesis.

13. A Remark to Wald's Paper: "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups," JUNJIRO OGAWA, University of North Carolina.

In the paper above mentioned, the late Professor A. Wald proposed a statistic U for the use in classification procedure and considered its exact sampling distribution. His result is too complicated to describe here. His proof was divided into nine lemmas, and each lemma was proved by an ingenious method. But, at least in the author's opinion, the proofs of his sixth and seventh lemmas can be improved with the help of the invariant measure defined on the Grassmann manifold, which consists of p-planes in $(n + 2)$-dimensional Euclidean space. The author presents new proofs of these lemmas as an example of applications of the theory of orthogonal group manifolds developed by A. T. James in 1954 (*Ann. Math. Stat.*, Vol. 25).

14. Consistency and Optimum Properties of Some Two-Sample Tests, JULIUS R. BLUM, Indiana University, and LIONEL WEISS, University of Virginia.

Let $X_1, \cdots, X_m$ be a sample from the uniform distribution on the unit interval and let $Y_1, \cdots, Y_n$ be a sample with density $g(y)$ on the unit interval. Let $Z_0 = 0$, Z_{n+1}, and $Z_1 < \cdots < Z_n$ be the order statistics corresponding to $Y_1, \cdots, Y_n$. For each $i = 1, \cdots, n + 1$ let S_i be the number of X's in the interval $[Z_{i-1}, Z_i]$, and for each non-negative integer r, let $Q_n(r)$ be the proportion among $S_1, \cdots, S_{n+1}$ which are equal to r. Let $\alpha = m/n$, and for each r, let $Q(r) = \alpha^r \int_0^1 \{g^2(y)/[\alpha + g(y)]^{r+1}\} \, dy$. Then it is shown that under mild restrictions on $g(y)$ we have $P\{\lim_{n\to\infty} \sup_{r\geq 0} | Q_n(r) - Q(r) | = 0\} = 1$. This is ap-

plied to prove consistency of certain two-sample tests such as the Wald-Wolfowitz run test (*Ann. Math. Stat.*, Vol. 11 (1940), pp. 147–162). One of these tests is shown to have a further desirable property.

15. Remarks on Characteristic Functions, Eugene Lukacs, The Catholic University of America and The Office of Naval Research.

Let $F(x)$ be a distribution function and denote by $\phi(t)$ its characteristic function (Fourier transform). Functions of characteristic functions are studied which are themselves characteristic functions. The following theorem is established: Let $\phi(t)$ be a characteristic function and let $G(z)$ be a function of the complex variable z which is analytic in $|z| < R$, where $R > 1$. The function $G[\phi(t)]$ is also a characteristic function if, and only if, $G(z)$ has a power series expansion about the origin with non-negative coefficients and if $G(1) = 1$. The class of functions $G(z)$ which have the property that $G[\phi(t)]$ is a characteristic function whenever $\phi(t)$ is a characteristic function includes also functions which are not analytic, for example the function $|z|^2$. By means of the theorem, one obtains also the following result: Let $\phi(t)$ be an arbitrary characteristic function and p be a real number such that $p > 1$; then, $(p-1)/[p - \phi(t)]$ is the characteristic function of an infinitely divisible distribution.

16. The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case, John S. White, University of Manitoba.

An auto-regressive process satisfying the stochastic difference equation $x_t = \alpha x_{t-1} + u_t$, $(t = 1, 2, \cdots)$, where the u's are independent identically distributed random variables, x_0 is a constant, and α is an unknown parameter, is said to be explosive if $|\alpha| \geqq 1$. If the u's are normally distributed with mean zero, it is shown that the maximum likelihood estimator for α has an asymptotic Cauchy distribution when $|\alpha| > 1$. For $|\alpha| = 1$, a characteristic function is obtained for the limiting distribution. For $\alpha = 1$, it is also shown that the limiting distribution of the maximum likelihood estimator for α is the distribution of a certain functional of a Wiener process.

17. The Distribution of the Ratio of Two Measures of Normal Dispersion, H. O. Hartley, Iowa State College.

Let us denote by x_i $(i = 1, 2, \cdots, n)$ a random sample of n items from $N(0, 1)$ and by $\bar{x}$ and s^2 the sample mean and variance, i.e., $\bar{x} = n^{-1}\Sigma x_i$; $s^2 = (n-1)^{-1}\Sigma(x_i - \bar{x})^2$. Consider now, the measure of dispersion $\phi = \phi(x_1 - \bar{x}, \cdots, x_n - \bar{x})$, where ϕ is a 1st order homogeneous function of its arguments $x_i - \bar{x}$, and finally $u = \phi/s$. Special cases of such a ratio which have been considered in the literature are: (a) ϕ = range = $x_{max} - x_{min}$ (David, Hartley and Pearson) *Biometrika*, 41, 482; (b) $\phi = x_{max} - \bar{x}$ (Pearson and Chandrasekar) *Biometrika* 28, 308; (c) $\phi = 1/n\Sigma \mid x_i - \bar{x} \mid$ (Geary) *Biometrika* 27, 310, 353. Here we develop a general distribution theory for the ratio u based on a fundamental integral equation: Introducing the probability integrals $F(U) = \Pr\{1/u \leqq U\}$; $G(U) = \Pr\{1/\phi \leqq U\}$ it follows from the independence of u and s that $G(U) = \int_0^\infty F(Us)\, f_\nu(s)\, ds$ where $f_\nu(s)$ is the ordinate distribution of s based on $\nu = n - 1$ degrees of freedom. Given the known integral $G(U)$ equation (1) is an integral equation Fredholm 1st kind for the unknown integral $F(U)$. Various methods of solving this equation are discussed and applied to cases (a) and (b), supplying answers unobtainable by the methods hitherto employed. Also, (d) $\phi = (\Sigma_{i=1}^{n-1} (x_i - x_{i+1})^2/(n-1))^{\frac{1}{2}}$ (von Neumann; *Ann. Math Stat.* 12, 367); (e) $\phi = \Sigma_{i=1}^{n-1} \mid x_i - x_{i+1} \mid/(n-1)$ (Kamat, *Biometrika*, 40, 116).

18. Estimating a Linear Functional Relation, H. FAIRFIELD SMITH, North Carolina State College.

The problem considered is to estimate a theoretical line $\xi_1 \cos\beta - \xi_2 \sin\beta - P = 0$ from paired observations $x_{1i}x_{2i}$ which are assumed to be random vectors from bivariate normal distributions around arbitrary centers $\xi_{1i}\,\xi_{2i}$. Most attempts to fit such a functional relation to observations with errors in both variates introduce the ξ_{pi} as what Neyman and Scott (1947) called "incidental parameters." These bring troubles to both least squares and maximum likelihood formulations. Attention is focused on the condition that the only ascertainable quantities from which a solution must be deduced are deviations of observations from the line in some prescribed direction. These have univariate distributions whose expectations in general deviate from the line by amounts proportional to their distances from the respective ξ_{pi}. But when, and only when, the deviations are measured in one particular direction their expectations are zero *independently of* ξ_{pi}. By utilizing this condition, and only thus, the incidental variables can be eliminated from the problem. The probability of a sample can then be expressed in terms of univariate normal distributions *about the line*, and maximum likelihood may be applied free of incidental variables. Kummell's solution is then seen to be unique and efficient. The estimator of the angle β is unbiased and its asymptotic variance may be evaluated. With certain supplementary conditions the exact sampling distribution has been obtained. (Supported by the Office of Ordnance Research).

19. Asymptotic Distributions of Roots of Certain Determinantal Equations, R. GNANADESIKAN, University of North Carolina.

The tests obtained by Roy for the hypotheses: (i) $\xi_1 = \xi_2 = \cdots = \xi_k$, i.e., $\Sigma^* = 0$ where Σ^* is the "between" covariance matrix, and (ii) $\Sigma_{12}(p \times q) = 0$ where Σ_{12} is the covariance matrix between a p-set and a q-set ($p \leq q$), on multivariate normal populations depend on the largest characteristic roots of (i) S^*S^{-1} where S^* and S are the sample "between" and "within" dispersion matrices respectively; and (ii) $S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}'$, where $S_{11}(p \times p)$ $S_{22}(q \times q)$ are sample covariance matrices of the p-set and the q-set respectively, and $S_{12}(p \times q)$ is the sample covariance matrix between the p-set and the q-set. The exact c.d.f. of this largest root has been obtained by Roy. For large sample sizes the problem becomes identical with that of finding the c.d.f. of the largest characteristic root of the sample dispersion matrix for a sample from one multivariate normal population. This limiting distribution has been obtained by Nanda for two particular cases, but there exists no explicit and general method of obtaining it. This has been done now. Also considering the test of, $\Sigma = \Sigma_0 = I(p)$ in particular, on one p-variate normal population we get Roy's test depending on the largest and the smallest characteristic roots of $S(p \times p)$, the sample covariance matrix. The joint c.d.f. of the largest and smallest roots has been obtained. Explicit expressions for some particular cases have also been obtained.

20. Investigation of the Possibility of Using Likelihood Ratio Tests of Certain Multivariate Hypotheses, for Obtaining Confidence Bounds, R. GNANADESIKAN, University of North Carolina, (By Title).

The likelihood ratio tests considered are of the following composite hypotheses on one or more multivariate normal populations $N[\xi(p \times 1), \Sigma(p \times p)]$: (i) $H_0: \Sigma = I(p)$ (one population), (ii) $H_0: \Sigma_1 = \Sigma_2$ (two populations), (iii) $H_0: \xi_1 = \cdots = \xi_k$ (analysis of variance of mean vectors for k populations) and (iv) $H_0: \Sigma_{12}(p \times q)(p \leq q) = 0$ (where Σ_{12} is the covariance matrix between a p-set and a q-set), the alternative in each case being $H \neq H_0$. One wants in each case confidence bounds (in terms of the observations) on meaningful

parametric functions which, as it were, would measure departures from the null hypothesis, such functions being (for the different cases) the respective characteristic roots of (i) Σ, (ii) $\Sigma_1\Sigma_2^{-1}$, (iii) $\Sigma^*\Sigma^{-1}$ (where Σ^* is the "between" and Σ the "within" dispersion matrix of the k populations) and (iv) $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$(where Σ_{11} and Σ_{22} are the dispersion matrices of the p-set and the q-set and Σ_{12} has been already defined). While the confidence bounds on these parametric functions are already available if one starts from other tests of these hypotheses and then inverts, it is found that if one starts from the likelihood ratio tests and then tries to invert, the problem of separation of the parametric functions from the observations becomes quite difficult.

21. Asymptotic Efficiencies of a Nonparametric Life Test for Investigating Smaller Percentiles of a Gamma Distribution, John E. Walsh, Lockheed Aircraft Corporation, (By Title).

In many life testing situations the quantity of interest is a specified smaller percentage point of the statistical population investigated. For example, a substantial loss may be incurred if more than a specified small percentage of the items of the population have the property of failing too soon. This paper considers some well known nonparametric tests of the sign test type and investigates their properties when applied to smaller percentage points for the case of a sample from a gamma probability distribution. Asymptotically, the nonparametric results are found to be highly efficient compared to the "best" parametric results based on the same fraction of items failed for the case of gamma distributions. Intuitive reasoning indicates that this high efficiency property holds for any reasonable type of statistical population and any sample size. Appropriate use of these nonparametric tests and estimates sometimes can yield a saving in cost and/or time without loss of statistical efficiency since the experiment can be stopped when only a fraction of the items being life tested are failed.

22. A Test of Judge Concordance for Paired Comparison Designs, (Preliminary Report), J. W. Wilkinson, University of North Carolina.

In a recent paper, (to appear in *Biometrika*), R. C. Bose has given several highly symmetrical designs where each of v judges compares r pairs of n objects, $(1 < r \leqq n(n-1)/2)$, and each pair is compared by k judges, $(1 < k \leqq v)$. To obtain a test of judge concordance for these designs, a pseudo-preference matrix $P_i = (p_{st}^i)$, $(s, t = 1, 2, \cdots, n)$, is constructed for each judge i, $(i = 1, 2, \cdots, v)$, where $p_{st}^i = 1$ or 0 according as a_s is, or is not, preferred to a_t. The diagonal cells, and cells p_{st}^i and p_{ts}^i when a_s and a_t are not compared by judge i, are left blank. A statistic Σ' is defined as $\Sigma' = \Sigma\gamma_{ij}(\gamma_{ij} - 1)/2$, where summation is extended over the non-diagonal cells of $P = \Sigma_{i=1}^{v} P_i$, and where γ_{ij} is the entry in cell (i, j) of P. The distribution of Σ' under the hypothesis that the preferences are allotted at random has been obtained, and has been tabulated for most of the known designs. Calculation of the first few moments of Σ' would indicate that a linear function of it is a χ^2 for large n and k. For $r = n(n-1)/2$, Σ' and its distribution are identical with those obtained by Kendall and B. Smith (*Biometrika*, Vol. 31, 1940).

23. On the Efficiency of Certain Classes of Tests Based on U-Statistics, Joan Raup Rosenblatt, National Bureau of Standards.

A class of non-parametric decision problems is characterized by partition of a set of possible probability distributions into sets defined by the value of a functional. Particular attention is given to a class of functionals of the form $E\phi(X; F)$, where X is a vector ran-

dom variable with distribution $F(x)$, and especially to the subclass in which the function $\phi(x)$ takes only the values zero and one. Certain families of tests are considered, which are based on functions of observed values of X which depend on these values only through the function $\phi(x)$. One such family is that based on the U-statistic corresponding to the functional $E\phi(X)$.

Methods are developed for computing an asymptotic expression for an index of efficiency for these families of tests relative to decision problems stated in terms of values of $E\phi(X; F)$. These methods are applied in particular to comparison of a family of two-sample tests based on the Wilcoxon-Mann-Whitney statistic with a family of tests based on a related statistic which has binomial distribution. Additional examples are given. (Work done at the Univ. of No. Carolina, with the support of the U. S. Air Force.)

24. The Dynamic Statistical Decision Problem when the Component Problem Involves a Finite Number, *m*, of Distributions, James F. Hannan, Michigan State University.

The dynamic problem consists in a sequence of N statistical decision problems with identical formal structure. Decisions are made successively within components and the risk of a sequence decision function is taken to be the average of the risks incurred in the components. An earlier paper by Hannan and Gaddum (submitted for an Annals of Mathematics Study) considered a sequence of formally identical two person S-games under the assumptions: (i) II's risk points in the component game form a closed and convex subset of the unit m cube, (ii) II's choice of strategy in each component can depend on the e.d. (empirical distribution) of I's moves in prior components. The principal theorem of that paper exhibited a usable sequence strategy (not depending on N) whose average risk across the N games exceeds by less than $(6m/N)^{\frac{1}{2}}$ the single-game minorant risk against the e.d. of I's N moves. The present paper is concerned with the substitution of estimates for the successive e.d. in (ii). If mixtures of the m distribution have unique representation, there exists a bounded vector kernel for unbiased estimates of the e.d. and the sequence decision function obtained by the substitution of unbiased estimates of this type satisfies the analogous theorem with the bound on the excess increased by multiplication by a bound of the kernel.

25. On Certain Systems of Experiments as Interdependent Stochastic Processes, (Preliminary Report), David Rosenblatt, American University, (By Title).

In certain systems of experiments, one may regard behavioral interaction between a "responsive" generalized subject and an experimenter in terms of interdependent discrete time-parameter stochastic processes. The subject (2) engages in actions or decisions A_{2k}; the experimenter (1), on the basis of an estimated conditional distribution of subject's actions, produces stimuli or treatments A_{1k}, where A_{ik} takes one of the values $a_{i1}, \cdots, a_{il}$, $i = 1, 2, k = 1, 2, \cdots$; subject and experimenter act alternately. The ith entity ($i = 1, 2$) possesses (a) a probability distribution R_{i0} over a finite set of m_i configuration states (threshold or preference configurations), i.e., a point in the m_i-dimensional simplex; (b) parameter stochastic operators ($m_i \times m_i$ matrices, row sum unity) Π_{ij}, $j = 1, \cdots, l$; (c) parameter mapping Γ_i or behavior function (matrix r.s.u.) which takes R_{ik} into the conditional distribution G_{ik} of A_{ik}, G_{ik} assuming values in the l-dimensional simplex. Let $f_j(A_{ik}) = 1$ if $A_{ik} = a_{ij}$, $= 0$ otherwise. Consider the system: (i) $R_{11} = R_{10}$; (ii) $R_{1.k+1} = R_{1k}(\Sigma_{j=1}^{l} f_j(A_{2k})\Pi_{1j})$; (iii) $R_{2k} = R_{2.k-1}\ (\Sigma_{j=1}^{l} f_j(A_{1k})\Pi_{2j})$; (iv) $G_{ik} = R_{ik}\Gamma_i$; $i = 1, 2$, $k = 1, 2, \cdots$, where each relation holds with probability one. The "conjoint process"

$E_k = (A_{2k}, R_{2k}, A_{1k}, R_{1k})$, $k = 2, 3, \cdots$, is a Markov process with stationary transition function. Conditions are adduced under which $\lim (G_{1k} - G_{2k}) = 0$, the null vector, with probability one, where G_{1k} is the experimenter's "estimate" of the "labile" conditional distribution of the subject's actions, G_{2k}. This model is constructed within the formal system OFK presented at the August, 1955, meeting of IMS (forthcoming abstract in *Econometrica*).

26. A Spherically-Symmetric Order Statistic r, (Preliminary Report), Brian Gluss and Fred L. Strodtbeck, University of Chicago.

n properties $A, B, \cdots, E$ are ordered $1, 2, \cdots, n$ by m people and scores $a, b, \cdots, e$ are calculated such that $q = \Sigma$ (ranks of A) $- [(n + 1)m/2]$, and so on. A statistic r is defined in the following manner: Consider n lines in an $(n - 1)$-dimensional space with orthogonal axes $(x_1, \cdots, x_{n-1})$, in which each line corresponds to one of the characteristics $A, B, \cdots, E$. The lines all pass through the origin and are such that the angles between all pairs are equal. Points $A, B, \cdots, E$ are defined on the lines such that the distances $OA, \cdots, OE$ (where O is the origin) are $a, b, \cdots, e$, respectively. A point $P(X_1 \cdots, X_{n-1})$ is then obtained such that OX_i is the sum of the projections of $OA, \cdots, OE$ on Ox_i. Then $OP = r$. The paper shows that $r = \sqrt{Sn/(n = 1)}$, where S is Kendall's statistic [M. G. Kendall, *Rank Correlation Methods*, 1948, Chap. 6.]. By this approach the paper shows it is possible by considerations of volumes of the space to test various hypotheses including (i) $A = B = \cdots = E$; (ii) $A > B > \cdots > E$ versus all other contingencies.

27. Generalized Normalization Polynomials, D. Teichroew, University of California at Los Angeles and National Cash Register Company, Dayton, Ohio.

Normalization polynomials have been studied by Campbell in 1923, Cornish and Fisher in 1937 and Hotelling and Frankel in 1938. Expansions based on these polynomials enable one to use the normal integral table for computing probability points of other distributions which approach the normal. These expansions have been expansions about the mean and the further the variate is from the mean the higher is the error of approximation. This paper shows that it is possible to get polynomials for expansions about an arbitrary point. These expansions are useful for obtaining probability points for probabilities very close to zero or one. Some expansions are obtained for the t and Gamma distributions. (Part of this work was sponsored by the Office of Naval Research).

28. "No Interaction" in a Three-way Table, Marvin A. Kastenbaum, University of North Carolina.

Let n_{ijk} denote the observed frequency, and p_{ijk} the probability of having an observation in the (ijk)th cell of a three-way table, $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s; k = 1, 2, \cdots, t)$. Also let the marginal probabilities $\Sigma_i p_{ijk} = p_{ojk}$, etc., $\Sigma_{i,k} p_{ijk} = \Sigma_k p_{ojk} = p_{ojo}$, etc., and $\Sigma_{i,j,k} p_{ijk} = 1$, and $\Sigma_{i,j,k} n_{ijk} = n$. Then, if the marginal frequencies are stochastic variates, the condition of independence between "(ij)" and "k" is expressed as: (1) $p_{ijk} = p_{ijo}p_{ook}$; and the conditions of independence between "i" and "k" and between "j" and "k" are given as (2) $p_{iok} = p_{ioo}p_{ook}$, and (3) $p_{ojk} = p_{ojo}p_{ook}$, respectively. A condition of "no interaction" is defined as one which, when taken together with (2) and (3), will yield (1). This condition is (4) $p_{ijk} = (q_{ijo}q_{iok}q_{ojk})/(q_{ioo}q_{ojo}q_{ook})$, where it is not assumed that $q_{ijo} = p_{ijo}$, etc., nor even that $q_{ioo} = \Sigma_j q_{ijo}$, etc. The q's in (4) may be eliminated in such a fashion as to yield distinct relationships among the p_{ijk}'s, namely: (5) $(p_{rst}p_{ijt})/(p_{ist}p_{rjt}) =$

$(p_{rsk}p_{ijk})/(p_{isk}p_{rjk})$, $i = 1, 2, \cdots, (r-1)$; $j = 1, 2, \cdots, (s-1)$; $k = 1, 2, \cdots, (t-1)$. If the hypothesis (4) of "no interaction" is to be tested, then the p_{ijk}'s may be estimated by maximizing the multinomial likelihood function, subject to (5) and to the further constraint $\Sigma_{i,j,k}\, p_{ijk} = 1$. The resulting $\hat{p}_{ijk}$'s, expressed in terms of their respective n_{ijk}'s and the deviations from expectation, are then substituted into the expression $\Sigma_{i,j,k}\, (n_{ijk} - n\hat{p}_{ijk})^2/n\hat{p}_{ijk}$, which for large n, is distributed as χ^2 with $(r-1)(s-1)(t-1)$ degrees of freedom.

29. On Bartlett's Test of Complex Contingency Table Interaction, SUJIT KUMAR MITRA, University of North Carolina.

Contrary to some of the current beliefs it is shown that the stochastic cubic equation suggested to Professor Bartlett (*J. Roy. Statist. Soc.*, suppl. 1935) by Professor R. A. Fisher in developing a test of his hypothesis of no interaction in a $2 \times 2 \times 2$ table might, with probability approaching one, have all the three roots real, under the null hypothesis of no-interaction. In such a case only the real root with the numerically smallest value will validate the use of χ^2 test. This could be immediately extended to a general $r \times s \times t$ table. A little thought would reveal that the numerical example considered by Bartlett actually represents 4 samples from 4 binomial populations. An attempt has been made to interpret interaction and main effects in this case and to furnish suitable tests for the suggested hypotheses.

30. A Theorem in Minimum Chi Square, SUJIT KUMAR MITRA, University of North Carolina, (By Title).

Let the possible results of a certain random experiment E be divided into r mutually exclusive groups and suppose that the probability of obtaining a result belonging to the ith group is $p_i^0 = p_i(\alpha_1^0, \cdots, \alpha_s^0)$ where $\alpha^0 = (\alpha_1^0, \cdots, \alpha_s^0)$ is an inner point of some non-generate interval A. We assume that $p_i(\alpha_1, \cdots, \alpha_s)$ considered as functions of $\alpha_1, \cdots, \alpha_s$ over A satisfy Cramér's conditions (a), (b), (c) and (d). (See Cramér: *Mathematical Methods of Statistics*, Section 30.3.) Let $f_k(\alpha_1, \cdots, \alpha_s)k = 1, 2, \cdots, t \leqq s$ be t functions of $\alpha_1, \cdots, \alpha_s$ such that for all points in A, the f_k satisfy the following conditions: (e) Every f_k has continuous derivatives $(\partial f_k)/(\partial \alpha_j)$ and $(\partial^2 f_k/(\partial \alpha_j \partial \alpha_l)$; (f) the matrix $\{(\partial f_k)/(\partial \alpha_j)\}$ where $k = 1, 2, \cdots, t$, $j = 1, 2, \cdots, s$ is of rank t. Let $f_k^0 (k = 1, 2, \cdots, t)$ be certain numbers in the range of the respective f_k's over A. We denote by H the hypothesis $f_k(\alpha_1^0, \cdots, \alpha_s^0) = f_k^0 (k = 1, 2, \cdots, t)$. Let ν_i denote the number of observations belonging to the ith group in n actual repetitions of E. Cramér has shown that the modified minimum chi-square equations have exactly one system of solutions $\hat{\alpha} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_s)$ such that $\hat{\alpha} \underset{P}{\longrightarrow} \alpha^0$ as $n \to \infty$ and such that $\chi_0^2 = \Sigma_{i=1}^r (\nu_i - np_i(\hat{\alpha}))^2/np_i(\hat{\alpha})$ is asymptotically distributed as a χ^2 with $r - s - 1$ d.f. The following results are proved: If H_0 is true (1) the modified minimum χ^2 equations subject to restriction H have exactly one system of solutions $\hat{\alpha}_H$ such that $\hat{\alpha}_H \underset{P}{\longrightarrow} \alpha^0$ as $n \to \infty$, and such that $\chi_H^2 = \Sigma_{i=1}^r (\nu_i - np_i(\hat{\alpha}_H))^2/(np_i\hat{\alpha}_H)$ is asymptotically distributed as χ^2 with $r - s + t - 1$ d.f.; (2) $\chi_H^2 - \chi_0^2$ is asymptotically independently distributed of χ_0^2 as a χ^2 with t d.f. This result is analogous to a result in least squares proved by C. R. Rao in his book. This also extends a result proved by Neyman (Proc. First Berk. Symp.).

31. Sequential Estimation from a Finite Population, HERBERT T. DAVID and INGRAM OLKIN, University of Chicago.

This paper is concerned with sequential estimation of the fraction defective p, in a finite (hypergeometric) population. The development is similar to that of Girshick,

Mosteller, Savage (*Ann. Math. Stat.*, v. 17, 1946, 13–23), who treat the case of infinite (binomial) populations. Path count ratios play essentially the same role in both cases, and are shown to provide unique unbiased estimates of certain functions of p when the regions are simple. Expressions for the variance of the estimate of p are given for both cases, and it is shown that for symmetric boundaries the variances in the finite and infinite situations are formally similar polynomials in pq of the same degree. A generalized finite population correction is discussed and, in particular, boundaries for which the variance is equal to cpq are considered.

32. **Tables for Computing Bivariate Normal Probabilities,** DONALD B. OWEN, Sandia Corporation.

A table of $T(h, a) = 1/(2\pi) \int_0^a \exp[-\frac{1}{2}h^2(1 + x^2)]/(1 + x^2)\, dx$, which may be used to obtain bivariate normal probabilities, has been computed to be used with a special two-dimensional linear interpolation scheme. The function is tabulated in two tables, one table having a coarse interval in one of the parameters and an interval fine enough for ordinary linear interpolation in the second parameter. The second table has the coarse interval on the second parameter and the fine interval on the first. By choosing the four points at the coarse intervals of the two tables that are nearest to a value to be interpolated and four other points on the fine intervals, the interpolation scheme gives accuracies comparable to ordinary linear interpolation with only ten per cent as many entries as that required for ordinary linear interpolation.

33. **Bounds and Approximations for Constants Used in Quality Control,** J. T. CHU, University of North Carolina and Case Institute of Technology.

Very close, yet very simple in form, upper and lower bounds are obtained for constants a, c_2, b, A, E_1, and B_i, $i = 1, \cdots, 4$, often used in quality control to set up control charts for individual observations and the means and standard deviations of groups of observations. For example, let a random sample: $x_1, \cdots, x_n$, be drawn from a normal distribution with variance σ^2. If $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n - 1)$ where $\bar{x} = \sum_{i=1}^n x_i/n$ and $a = E(s)/\sigma$, then $[(2n - 3)/(2n - 2)]^{\frac{1}{2}} \leq a \leq [(2n - 2)/(2n - 1)]^{\frac{1}{2}}$ for all integers $n \geq 2$. In using these bounds and their arithmetic mean as approximations to a, the proportional errors are shown to be respectively less than $E = \frac{1}{2}(2n - 1)(2n - 3)$ and $E/2(E/2 = .004$ if $n = 5)$. Similar results are obtained for the constants mentioned above. Tables are given for illustration (Research partially supported by the Office of Naval Research).

34. **Four Streams of Traffic Converging on a Cross-Road,** BRIAN GLUSS, University of Chicago, (By Title).

Four streams of traffic arrive at a cross-road in independent Poisson process. The lights operate such that if they have just turned red against two of the streams they will turn green again when either (i) n cars are waiting, or (ii) a time α has passed, whichever is the sooner. n and α are prearranged. The frequency function and expectation of the waiting time τ of a car wishing to go straight or to turn right are obtained: $E(\tau) = [n(n + 1)/2 - e^{-m\alpha}\sum_{R=0}^n (n - R + 1)(n + R)(m\alpha)^R/2(R)!]/(t_1 + t_2)m^2$, where m = sum of mean flows of the two streams for which the lights are red, and t_1 and t_2 are the expected red time-periods for the two sets of two streams. The frequency functions of the waiting times u, v of a car arriving in a green or red period respectively and wishing to turn left, and therefore having

to wait for a sufficient time-gap in the opposing stream or until the lights turn red again, are found. $E(u)$ and $E(v)$ are then calculated for some sets of parameter-values by approximate integration.

35. **Markov Processes Arising in Learning Models,** John G. Kemeny and J. L. Snell, Dartmouth College.

The paper studies two learning models, one due to W. K. Estes, and the other due to R. R. Bush and F. Mosteller. In the cases studied both models lead to one-parameter families of Markov processes; the Estes model having a finite number of states, the Bush-Mosteller model an infinite number. For each value of the learning-parameter there is a single Bush-Mosteller process, but an infinite number of Estes models—one for each possible number of states. It is shown that for a given value of the learning-parameter, as the number of states tends to infinity, the stationary distribution of the Estes processes tends to the stationary distribution of the corresponding Bush-Mosteller process. Moments of the stationary distribution are found in the Bush-Mosteller processes, and the distributions themselves are also found in several special classes of processes. It is shown that as the learning-parameter tends to zero the stationary distributions in both models approach very simple distributions. Since some psychologists are interested primarily in low values of the learning-parameter, this result provides simple approximate answers.

36. **On a Decision Rule for Selecting a Group Containing the Population with the Largest Mean, (Preliminary Report),** R. C. Bose and S. S. Gupta, University of North Carolina, (By Title).

Suppose there are $(n + 1)$ normal populations $N(\mu_i, \sigma^{*2})$, $i = 0, 1, 2, \cdots, n$, and that $x_0, x_1, \cdots, x_n$ are the $(n + 1)$ means based on samples of equal size k, one from each population. One would like to select as small a group as possible subject to the restriction that the least upper bound of the probability of not including in the group the population with the largest mean is $\alpha(0 < \alpha < 1)$. K. C. Seal (*Ann. Math. Stat.*, Sept., 1955) has given an infinite class of decision rules for this problem and has obtained an optimum rule for the situation when all means but one are equal. Another rule has been studied here in detail. This is based on the auxiliary statistic $u = (Y_{(n)} - Y_0)/s$, where $Y_0, Y_1, \cdots, Y_n$ are independently and identically distributed $N(0, \sigma^2)$, $\sigma^2 = \sigma^{*2}/k$, corresponding random observations $y_1, y_2, \cdots, y_n$ being ordered as $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$ and where s^2 is an independent estimate of σ^2. The rule is, "Reject any observation x_0 from the given $x_i (i = 0, 1, 2, \cdots, n)$ if $x_{(n)} - x_0 > su_\alpha$ and retain otherwise; where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are n ordered observations among $(x_1, x_2, \cdots, x_n)$ and u_α is the upper α % point of $(Y_{(n)} - Y_0)/s$." The upper 5 % points of U, in the case when σ is known, have been tabulated.

37. **Recurrent Values of Sums of Independent Random Variables, (Preliminary Report),** Louis J. Cote and Henry Teicher, Purdue University.

Let $\{X_i\}$ $i = 1, 2, \cdots$ be a sequence of independent random variables defined on a probability space with $S_{m,n} = \sum_{i=m}^{n} X_i$, $S_{1,n} = S_n$ and $P_\epsilon = P\{| S_{m,n} - b | < \epsilon \text{ i.o.}\}$ where i.o. signifies "infinitely often" (i.e., for infinitely many values of $n \geqq m$). The real number b is called recurrent or quasi recurrent for the sequence $\{S_{m,n}\}$ according as $P_\epsilon = 1$ or $P_\epsilon > 0$ for all $\epsilon > 0$. The classes of such values are examined and conditions for the existence of recurrent or quasi-recurrent values are considered.

38. A Problem Involving the Distribution of Shadows, (Preliminary Report), HERMAN CHERNOFF, Stanford University, and JOSEPH F. DALY, National Bureau of the Census.

A source of light is at a point P and a worm is crawling in a given direction along a line L which does not go through P. Circular disks are distributed randomly throughout the plane containing P and L. Suppose that the worm can travel only in the shadow. The distribution of the distance the worm can travel from a given starting point is characterized. One such characterization involves the wave equation. The results generalize to the cases where the disks are replaced by line segments parallel to L and the source of light is at infinity and these results have applications to geiger counter and traffic problems. The corresponding problems of the worm who travels only in light are rather easy to treat.

39. Note on Two-stage Test Procedures, S. G. GHURYE, Lucknow University, (By Title).

This note concerns tests of hypotheses regarding a parameter which are designed to have power independent of another parameter. The conditions satisfied in the problem of the mean of a normal distribution solved by Stein (*Ann. Math. Stat.*, 1945) are stated more generally, and the corresponding general solution is given. It is shown that these conditions are also satisfied in the problem of testing for the location parameter of an exponential distribution by a number of two-stage tests, and the performances of some of these tests are compared in some particular cases.

40. Some Properties of Generalized Sequential Probability Ratio Tests, JACK C. KIEFER, Cornell University, and LIONEL WEISS, University of Virginia.

Generalized sequential probability ratio tests (GSPRT) are known to form a complete class with respect to the probabilities of making errors and the distribution of the sample size, when one simple hypothesis is being tested against another. In this paper it is shown that (1) under certain conditions, a GSPRT is uniquely determined by the distributions of the sample size under the two hypotheses; (2) for a GSPRT to be admissible with respect to the probabilities of error and the distribution of the sample size, the decision bounds characterizing the test must obey certain inequalities; (3) under certain monotonicity conditions on the probability ratio, a GSPRT forms a complete class with respect to the probabilities of error and the "average" distribution of the sample size (averaged over a set of alternatives to the two hypotheses being tested); and (4) a class of tests complete with respect to the probabilities of error and the expected sample size under a third distribution consists of truncated GSPRT whose decision bounds satisfy certain inequalities.

41. Sequential Decision Problems for a Class of Stochastic Processes. Testing Hypotheses, (Preliminary Report), A. T. BHARUCHA-REID, University of California, Berkeley.

Let $\{X_i(t), t \geqq 0\}$, $i = 1, 2$, be two different stochastic processes with continuous time parameter. Beginning at $t = 0$, a process $\{X(t), t \geqq 0\}$, which is either $\{X_1(t)\}$ or $\{X_2(t)\}$, is observed continuously, and on the basis of the observed realization of the process, $x(t)$, the statistician wishes to decide whether $\{X(t)\}$ is $\{X_1(t)\}$ or $\{X_2(t)\}$. This problem has been considered by Dvoretzky, Kiefer, and Wolfowitz (*Ann. Math. Stat.*, Vol. 24 (1953), pp. 254–264) when $\{X(t)\}$ is a stochastic process with stationary independent increments.

In this paper we consider the case where $\{X(t)\}$ is a Markov process, with $x(t)$ a sufficient statistic for the process. We consider in particular the application of these results to some branching stochastic processes, e.g., the birth, death, birth-and-death, and Pólya processes. Let $p(x, t; \omega) = \Pr(X(t) = x \mid \omega)$, $x = 0, 1, 2, \cdots$; $\omega \in \Omega$, and denote by $D(t)$ the decision function $\log \{p(x, t; \omega_2)/p(x, t; \omega_1)\}$. For decision boundaries A and B, $B < 0 < A$, the Wald sequential procedure is used to test the hypothesis $H_i (i = 1, 2)$ that $\omega = \omega_i$, where ω_1 and ω_2 are any two positive numbers, $\omega_1 \neq \omega_2$. Let $f(d; \omega)$ denote the probability that the decision procedure will terminate with the acceptance of H_2 when the parameter is really ω and $D(0) = d$; and let $m(s, \tau) = E\{\exp(s\tau)\}$ be the moment generating function of the observation time τ necessary to reach a decision when $D(0) = d$ and the parameter is really ω. The usual probabilistic reasoning leads to functional equations for $f(d; \omega)$ and $m(s, \tau)$, the analytic properties of which will be discussed in a subsequent publication. (This work was supported by the USAF School of Aviation Medicine.)

42. **Note on a Markov Chain with Matrix States and Some Applications,** A. T. Bharucha-Reid and Rodabé P. Bharucha-Reid, University of California, Berkeley, (By Title).

In connection with a probability problem in learning theory concerned with latent and reinforced types, it was necessary to consider a Markov chain with matrix states. Various ways of defining the transition probabilities are considered, and the asymptotic properties of the chain investigated. The results obtained are applicable to the study of changes in systems whose structure has a matrix representation, e.g., communication nets, social groups, etc.

43. **On the Comparison of Two Stochastic Epidemics,** A. T. Bharucha-Reid, University of California, Berkeley, (By Title).

In this paper the Girshick procedure for comparing or ranking two populations with respect to an unknown parameter is applied to the problem of comparing the effect of two types of housing on hospital admission rates for acute respiratory disease. The procedure is applied when different stochastic models are used to describe the development of the epidemic. Data used are from an epidemic situation studied at Sampson Air Force Base, Geneva, New York. (This work was supported by the USAF School of Aviation Medicine.)

44. **A Sequential Multiple Decision Procedure for Selecting the Population with the Largest Mean from k Normal Populations with a Common Unknown Variance, (Preliminary Report),** R. E. Bechhofer, Cornell University, and M. Sobel, Bell Telephone Laboratories, (By Title).

Let $x_{ij} (i = 1, \cdots, k; j = 1, 2, \cdots)$ be independent observations from normal populations Π_i with unknowns means μ_i and a common unknown variance, and let $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$ denote the ranked means. A sequential procedure is proposed which guarantees probability P^* of selecting the population with the largest mean $\mu_{[k]}$ whenever $\mu_{[k]} - \mu_{[k-1]} \geqq \delta^*$; the constants $P^* < 1$ and $\delta^* > 0$ are preassigned. Let

$$\bar{x}_{im} = \textstyle\sum_{j=1}^{m} x_{ij}/m, \qquad s_m^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} (x_{ij} - \bar{x}_{im})^2/k(m-1),$$

and $t_{ijm} = (\bar{x}_{im} - \bar{x}_{jm} - \delta^*)/s_m\sqrt{2/m}$. For $i = 1, \cdots, k$ let

$$L_{im} = \left[1 + \sum_{\substack{\alpha=1 \\ \alpha,\beta \neq i}}^{k} \sum_{\beta=1}^{k} A_{\alpha\beta}\, t_{i\alpha m}\, t_{i\beta m}/k(m-1)\right]^{-(km-1)/2},$$

where $A_{\alpha\beta} = 2(k-1)/k$ for $\alpha = \beta$ and $-2/k$ for $\alpha \neq \beta$, let $L_{[1]m} \leqq \cdots \leqq L_{[k]m}$ denote the ranked L_{im}, and let $P_m = L_{[k]m}/\Sigma_{i=1}^{k} L_{im}$. At every stage, the k values, L_{im} differ with probability one and are in one-to-one correspondence with the k populations Π_i; let $\Pi_{[k]m}$ denote the population associated with $L_{[k]m}$ at the mth stage. *Procedure*: At the mth stage $(m = 1, 2, \cdots)$ take the vector observation $(x_{1m}, \cdots, x_{km})$ and compute P_m. If $P_m \geqq P^*$, stop and select $\Pi_{[k]m}$; if $P_m < P^*$, take the $(m+1)$st vector observation and compute P_{m+1}. This procedure meets the requirement, is scale and location invariant, and the probability of termination is unity. The procedure can be generalized to handle problems such as obtaining a complete ranking of the k means. (Research supported in part by the U. S. Air Force through the Office of Scientific Research of the ARDC.)

45. A Scale Invariant Sequential Multiple Decision Procedure for Selecting the Population with the Smallest Variance from k Normal Populations, (Preliminary Report), R. E. Bechhofer, Cornell University, and M. Sobel, Bell Telephone Laboratories, (By Title).

Let $x_{ij} (i = 1, \cdots, k; j = 1, 2, \cdots)$ be independent observations from normal populations Π_i with unknown means μ_i and unknown variances σ_i^2, and let $\sigma_{[1]}^2 \leqq \cdots \leqq \sigma_{[k]}^2$ denote the ranked variances. A sequential procedure is proposed which guarantees probability P^* of selecting the populations with the smallest variance $\sigma_{[1]}^2$ whenever $\sigma_{[2]}^2/\sigma_{[1]}^2 \geqq \theta^*$; the constants $P^* < 1$ and $\theta^* > 1$ are preassigned. Let $\bar{x}_{im} = \Sigma_{j=1}^{m} x_{ij}/m$, $s_{im}^2 = \Sigma_{j=1}^{m} (x_{ij} - \bar{x}_{im})^2$, and $R_{jim} = s_{jm}^2/s_{im}^2$. For $i = 1, \cdots, k$ let

$$L_{im} = [\Pi_{j=1}^{k} R_{jim}]^{(m-3)/2}[1 + \Sigma_{j=1, j\neq i}^{k} R_{jim}/\theta^*]^{-k(m-1)/2},$$

let $L_{[1]m} \leqq \cdots \leqq L_{[k]m}$ denote the ranked L_{im}, and let $P_m = L_{[k]m}/\Sigma_{i=1}^{k} L_{im}$. At every stage, the k values, L_{im}, differ with probability one and are in one-to-one correspondence with the k populations Π_i; let $\Pi_{[k]m}$ denote the population associated with $L_{[k]m}$ at the mth stage. *Procedure*: At the mth stage $(m = 2, 3, \cdots)$, take the vector observation $(x_{1m}, \cdots, x_{km})$ and compute P_m. If $P_m \geqq P^*$, stop and select $\Pi_{[k]m}$; if $P_m < P^*$ take the $(m+1)$st vector observation and compute P_{m+1}. This procedure meets the requirement, is scale and location invariant, and the probability of termination is unity. The procedure can be generalized to handle problems such as obtaining a complete ranking of the k variances. A similar procedure can be used for the case of known means and for ranking the scale parameters of exponential populations. (Research supported in part by the U.S. Air Force through the Office of Scientific Research of the ARDC.)

46. Exact Probabilities in a Test for Markoff Dependency, Reed B. Dawson, Jr., Department of Defense.

This paper is concerned with Markoff dependency (of the first order) in a digital stream, where the object is to test the hypothesis of independence against any alternative which alters the probabilities of the pairs. Let N digits be distributed about an oriented circle so that each of the $(N-1)!$ arrays are equally likely, and form a matrix $[f_{ij}]$, where f_{ij} is the number of digits i which are followed by a digit j. The exact probability of this matrix of pairs is found, generalizing a result of Stevens (*Ann. Eugenics*, Vol. 9 (1939), pp. 10–17). This probability, asymptotically the same as the probability that a matrix with the same entries will arise from the usual contingency table assumptions, illuminates a special case of the asymptotic test of Hoel (*Biometrika*, Vol. 41 (1954), pp. 430–433) for Markoff dependency of general order. A formula for the expectancy of a product of factorial powers of the f_{ij} is derived.

47. A Combinatorial Problem and Its Application to Probability Theory, T. V. NARAYANA, McGill University.

A quasi order called k-domination is defined on the r-partitions of two integers m and n. An explicit expression for the number of k-dominations of the r-partitions of n by those of m is derived. This result is extended and shown to be a generalization of the "problème du scrutin" of D. André. Two classes of coin-tossing problems are solved as an application of this result. A number of combinatorial identities and the solution of a class of difference equations are obtained by probability methods. The relation of this problem to the recurrent events of Feller in the case of coin tossing is briefly discussed.

48. The Bayesian Inference Problem in Stochastic Systems, MAX A. WOODBURY, George Washington University.

In an experimental or environmental stochastic system, the possible inputs to the controlled stochastic process are represented by stochastic mappings of the internal states of the system into each other. The observable outputs are assumed to be the result of a stochastic mapping from the internal states of the system to the set of possible outputs. In the case where the inputs only are known, the general formula for the a posteriori distribution at a given time is the result of applying the product of the input stochastic matrices to the a priori distribution vector. If, however, account is taken of the information provided by the output the result is expressible in linear terms only if the requirement for a normalized probability vector is dropped. The relationship of this result to the stochastic behavior models of Rosenblatt, Flood and Mosteller is discussed. (The research covered by this abstract was supported by the Office of Naval Research.)

49. Some Nonparametric Generalizations of Multivariate Analysis and Analysis of Variance, S. N. ROY, University of North Carolina.

With observed frequency data arranged in a multi-way table, assuming that the observations are independent in probability, there will be, under any hypothesis, (i) a single multinomial distribution or (ii) the product of a number of separate multinomial distributions according as (i) only the total number of observations is supposed to be fixed from sample to sample or (ii) marginal frequencies in certain directions of the table are supposed to be fixed. An attempt is made at a systematic elaboration of the historically prior ideas of Barnard and Pearson (*Biometrika*, 1947), to (i) multivariate analysis, starting from a single multinomial and framing hypothesis suitable to multivariate analysis situations and to (ii) analysis of variance, starting from the product of an appropriate number of multinomials and framing hypotheses suitable to analysis of variance situations. The theorems used are those of Cramér [*Mathematical Methods of Statistics*, Chapter 30] and some other theorems which can be proved on the same lines. The conditional probability approach is altogether abandoned.

50. Further Remarks on Measures of Association for Cross-Classifications, LEO A. GOODMAN, University of Chicago, and WILLIAM H. KRUSKAL, Universities of California and Chicago.

Measures of association discussed by the authors previously (*J. Amer. Stat. Assn.*, 49 (1954), 732–64) are considered further, especially in regard to the sampling distributions of their sample analogues. Asymptotic distributions are obtained for a number of cases,

and numerical investigations of the accuracy (qua approximations) of these asymptotic distributions are described.

51. Uniformly Consistent Sequences of Multiple-Decision Rules, William Jackson Hall, University of North Carolina.

Suppose x has an unknown distribution function F, belonging to one of m disjoint classes $\omega_1, \cdots, \omega_m$, and suppose $A_1, \cdots, A_m$ are corresponding alternative decisions, one of which is to be chosen by a multiple-decision rule (m-d.r.) D_n after taking a sample of size n. D_n is defined by $\phi^n(x) = [\phi_1^n(x), \cdots, \phi_m^n(x)]$, $\phi_i^n \geqq 0$, x denoting the sample, where the θ_i^n's sum over i to unity. $\phi_i^n(x)$ is the probability that D_n chooses A_i when x is observed. *Definition* 1: $\{\phi^n(x)\}$, $n = 1, 2, \cdots$, defines a "uniformly consistent sequence (u.c.s.) of m-d.r.'s $\{D_n\}$ for discriminating among $\omega_1, \cdots, \omega_m$" if $\lim_{n\to\infty} \inf_{F\in\omega_i} \mathcal{E}_F\phi_i^n(X) = 1$ $(i = 1, \cdots, m)$. *Definition* 2: $\phi^n(x)$ defines a "non-trivial m-d.r. D_n for discriminating among $\omega_1, \cdots, \omega_m$" if $\sum_{i=1}^m \inf_{F\in\omega_i} \mathcal{E}_F\phi_i^n(X) > 1$. *Theorem* 1: *A necessary and sufficient condition for the existence of a* u.c.s. *of* m-d.r.'s *for discriminating among* $\omega_1, \cdots, \omega_m$ *is that there exist non-trivial* 2-d.r.'s *for discriminating between* ω_i *and* ω_j *for some* n_{ij} *(sample size) for every* $i \neq j$. Results of Hoeffding (unpublished) and Berger and Wald (*Ann. Math. Stat.*, Vol. 20 (1949), pp. 104–9) are adapted to supply some necessary and sufficient conditions, respectively, for the existence of non-trivial 2-d.r.'s. *Theorem* 2: *A necessary and sufficient condition for the existence of a most economical* m-d.r. *relative to any* $(\alpha_1, \cdots, \alpha_m)$, *or relative to any* (β_{ij}), *for discriminating among* $\omega_1, \cdots, \omega_m$ (Hall, Abstract, *Ann. Math. Stat.*, Vol. 25 (1954), p. 814) *is that there exist a* u.c.s. *of* m-d.r.'s *for discriminating among* $\omega_1, \cdots, \omega_m$.

52. Some Hypergeometric Series Distributions Occurring in Birth-and-Death Processes at Equilibrium, (Preliminary Report), William Jackson Hall, University of North Carolina, (By Title).

Some time-homogeneous birth-and-death processes at equilibrium are considered in which the birth and death rates are "stimulated" by "overcrowding." Generally, under mild restrictions, p_n, the distribution of population size n is proportional to $\Lambda_n/(M_n n!)$, where $\Lambda_n = \lambda_0\lambda_1 \cdots \lambda_{n-1}$ and $M_n = \mu_1\mu_2 \cdots \mu_n$; $\lambda_n(\mu_n)$ is the birth (death) rate when the population size is n. If λ_n is quadratic in n (i.e., constant immigration rate and reproduction rate is linear in n) and μ_n linear in n, then p_n is shown to be proportional to the $(n + 1)$th term in a general hypergeometric series, a four parameter distribution. If λ_n and μ_n are both linear (constant reproduction rate), p_n is proportional to the $(n + 1)$th term in a confluent hypergeometric series, a three parameter distribution. In the same manner, using a constant death rate, p_n is proportional to the $(n + 1)$th term in a negative binomial series, as is well known; and, with no reproduction, an exponential series (Poisson distribution) is obtained. Each distribution is a limiting form of the preceding one. Generating functions, moments, and approximate estimates by the method of moments of the parameters of the hypergeometric series distributions are derived.

53. Some General Aspects of Stochastic Approximations, Tosio Kitagawa, Iowa State College.

As one continuation of random integration introduced by the author, some general aspects of stochastic approximations will be discussed specifically in reference to the risk function formulations. Our stochastic approximations are concerned with the various problems of (a) solutions of equations, (b) interpolation problems, (c) mapping problems, and (d) numerical differentiations.

54. The Analysis of Incidence Rates Under Multiple Classifications of the Population, (Preliminary Report), Wyman Richardson, University of North Carolina.

A population is classified two ways into cells, n_{ij}. The number of cases, a_{ij} (of some disease, for instance), is assumed to have, in one model, a Poisson distribution with parameter $n_{ij}\rho_{ij}$, and in another, a binomial (Q_{ij}, n_{ij}) distribution. Q_{ij} is assumed to be equal to $f(\theta_i, \psi_j)$. The hypothesis $\psi_1 = \cdots = \psi_s$ can be tested in each model by χ^2, with expected frequencies in each cell of $N_{ij}A_{i\cdot}/N_{i\cdot}$, (where $A_{i\cdot} = \Sigma_j A_{ij}$, etc.). Maximum likelihood equations are derived for the case $f(\theta_i, \psi_j) = \theta_i\psi_j$. It is shown that, except for a multiplicative factor, there is a single solution of these equations, which can be obtained by efficient iterative procedures. This result holds when there are k classifications. In the Poisson model, these estimates are sufficient. A large sample test of the hypothesis $\psi_1 = \cdots = \psi_s$ against the alternative $Q_{ij} = \theta_i\psi_j$ is to compute $\chi^2 = 2[\Sigma A_{i\cdot} \log_e (\hat{\theta}_i N_{i\cdot}/A_{i\cdot}) + \Sigma A_j \log_e \hat{\psi}_j]$.

55. Estimation of Percentiles by Order Statistics, A. E. Sarhan, University of North Carolina.

In previous work of the author ("Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), and Part II, *Ann. Math. Stat.*, Vol. 26 (1955)) the means and standard deviations of certain distributions were estimated by the best linear combinations of the ordered sample values. In the present paper, the same methods are used to derive a general expression for estimation of the jth percentile and its variance. From this expression and by making use of previous results, the jth percentile is estimated for certain distributions. As special cases, the estimates of the 50th percentile (the population median) and of the semi-interquartile range are calculated.

56. On Renewal Theory, Counter Problems, and Quasi-Poisson Processes, Walter L. Smith, University of North Carolina.

Let $\{t_i\}$ be a *renewal process*, i.e., a sequence of non-negative, independent, identically distributed random variables which are not zero with probability one. Let $\mu_r = Et_i^r$, and define n_t by $\Sigma_1^{n_t} t_i \leq t < \Sigma_1^{n_t+1} t_i$ (taking $n_t = 0$ if $t_1 > t$). If $H(t) = En_t$, then it is shown that (i) if $\mu_1 < \infty$, then a necessary and sufficient condition for $\mu_2 < \infty$ is that $\lim_{t\to\infty} \{H(t) - t\mu_1^{-1}\} = \beta$ exist and be finite, when $\mu_2 = 2\mu_1^2(1 + \beta)$; (ii) if $t_i = u_i + v_i$ where $\{u_i\}$ and $\{v_i\}$ are independent renewal processes, the v_i having a negative exponential distribution, then $\lim_{t\to\infty} H'(t)$ always exists. The results (i) and (ii) render the calculation of the asymptotic properties of a certain electronic counter process, previously studied by Hammersley, straightforward. If $H(t)$ is linear in t for all $t > \tau$, for some finite τ, the process is called quasi-Poisson, and the class of quasi-Poisson processes is not empty. Let $Y(x; t) = \Pr(\Sigma_1^{n_t+1} t_i \leq t + x)$. Then it is shown that a necessary and sufficient condition for $Y(x; t)$ to be independent of $t > \tau$ is that $\{t_i\}$ be quasi-Poisson. When $\{t_i\}$ is quasi-Poisson, $\mu_r < \infty$ for all r, and the study of the effects of an automatic self-paralyzing mechanism on $\{t_i\}$, of a type in use for blood-cell counting, becomes trivial.

57. On the Construction of Significance Tests on the Circle and the Sphere, G. S. Watson, The Australian National University, and E. J. Williams, Commonwealth Scientific and Ind. Res. Organization, S. Melbourne, (By Title).

The probability density proportional to exp $(k \cos \theta)$, where k is a precision constant and θ is the angle between an observed vector and a population mean vector or polar vector, has

been considered in two and three dimensions by several authors. Significance tests are required to test (i) that $k = k0$, is a prescribed value, or that several populations have the same value of k, and (ii) that the polar direction of a population has prescribed direction cosines or that several populations have the same polar vectors. Tests of these hypotheses are given which are free of nuisance parameters. They are based on conditional distributions formed by holding constant sufficient statistics. Inequalities and approximations are suggested to make the tests easy to apply in practice. The arithmetic examples given suggest that, in three dimensions, the tests given by one of us (G.S.W.) elsewhere will be satisfactory.

58. **Estimation of Individual Variations in an Unreplicated Two-Way Classification, (Preliminary Report),** THOMAS S. RUSSELL and RALPH ALLAN BRADLEY, Virginia Polytechnic Institute, (By Title).

Consider a two-way classification, the usual model $x_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, $i = 1, \cdots, n$, $j = 1, \cdots, r$ (e.g., r chemists and n batches or r judges and n items) and the usual assumptions, except let var $(\epsilon_{ij}) = \sigma_j^2$. It was assumed that an estimator Q_j of σ_j^2 should be a quadratic form in the $(r - 1)(n - 1)$ linear contrasts usually ascribed to error. Reasonable requirements on such a quadratic form led to the estimator in $(n - 1)(r - 1)(r - 2)Q_j = r(r - 1)E_j - E$, where $E_j = \sum_{i=1}^n (x_{ij} - x_{i.} - x_{.j} + x_{..})^2$ and $E = \sum_{j=1}^r E_j$, the usual error sum of squares. Q_j is the estimator previously suggested by Ehrenberg (*Biometrika*, Vol. 37 (1950), pp. 347–357). Q_j has been shown to be the maximum likelihood estimator of σ_j^2 only when $r = 3$. When $\sigma_j^2 = \sigma^2$ for all j, the distribution of Q_j/σ^2 has been shown to be that of $r\chi^2_{n-1}/(n - 1)(r - 1) - \chi^2_{(n-1)(r-2)}/(n - 1)(r - 1)(r - 2)$, the two χ^2's being independent. Q_j/E has been shown to be a monotone function of an F with $(n - 1)(r - 2)$ and $(n - 1)$ degrees of freedom formed from those χ^2's. The joint distribution of the Q's has been considered and further research on various aspects of the problem is underway. (Work supported by A.R.S., U.S.D.A. and Q.M.R. and D., U.S. Army.)

59. **Empirical Bayes Estimation, (Preliminary Report),** M. V. JOHNS, JR., Columbia University.

Let $\mathbf{X} = (X_1, X_2, \cdots, X_r)$ where the X_i's are independent discrete valued random variables with a common c.d.f. $F(x \mid \lambda)$, and where there exists an a priori probability measure p over a σ-algebra of subsets of the values of the parameter λ, so that the parameter is also a random variable Λ. Suppose that it is desired to estimate $\theta(\lambda) = E(X_i \mid \Lambda = \lambda)$, using the risk function $E(\varphi(\mathbf{X}) - \theta(\Lambda))^2$, where $\varphi(\mathbf{x})$ is any estimator. Let the Bayes estimator (depending on p and $F(x \mid \lambda)$) be $\varphi^*(\mathbf{x}) = E(\theta(\Lambda) \mid \mathbf{X} = \mathbf{x})$. Suppose now that p and the form of F are unknown, but that n independent $(r + 1)$-component random vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$, each having the same probability structure as $\mathbf{X}$, are available. Then a "non-parametric" estimator $\varphi_n(\mathbf{x}) = \varphi_n(\mathbf{x}; \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ is given having the property that $\lim_{n\to\infty} E(\varphi_n(\mathbf{X}) - \theta(\Lambda))^2 = E(\varphi^*(\mathbf{X}) - \theta(\Lambda))^2$ for any p and F subject to certain mild restrictions. The general case where the X_i's are not necessarily discrete valued is also considered. Similar results are obtained for several cases (considered by Robbins in the *Third Berkeley Symposium on Mathematical Statistics and Probability*) where p is unknown but F belongs to a specified one-parameter family of probability distributions and where the value of the parameter is to be estimated. The behavior of these empirical Bayes estimators is also investigated for finite n for certain special cases.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

T. A. Bancroft, director of the Statistical Laboratory, Iowa State College, has been in Mexico from midAugust to early October on behalf of the Food and Agriculture Organization, UN. His assignment was to advise the Government of Mexico and FAO on the present status of the use of experimental designs and survey techniques in Mexico and to help prepare for the 1956 FAO-sponsored training center in Mexico.

R. E. Beckwith has left the Operations Research Group, Case Institute of Technology, to join the staff of the Statistical Laboratory, Purdue University, Lafayette, Indiana.

Dr. Charles B. Bell, Jr., formerly research engineer at Douglas Aircraft Company, has been appointed Assistant Professor of Physics at Xavier University, New Orleans, Louisiana.

Dr. Ernest P. Billeter has left the Statistical Office of the City of Zurich to become a Scientific Assistant to the Bank for International Settlements (BIS) in Basel.

Benjamin Buchbinder is now a statistician in the Commissioned Corps of the U. S. Public Health Service, assigned to the Division of Nursing Resources of the Department of Health, Education and Welfare in Washington, D. C.

Donald L. Burkholder, who received his Ph.D. in mathematical statistics at the University of North Carolina in 1955, is now an Assistant Professor in the Department of Mathematics of the University of Illinois.

Theodore Colton is now a statistician in the Commissioned Corps of the U. S. Public Health Service in the Program Evaluation and Reports Branch of the Division of Hospital and Medical Facilities.

John H. Cover has joined the HRAF South Asia Project at the University of California until June 30, 1956, at which time he will return to his position with the Bureau of Business and Economic Research at the University of Maryland.

Don A. Davis of Seattle, Washington, is temporarily located at 145-C Baker Village, Columbus, Georgia, where he is with the Army at Fort Benning.

Claude de Courval is now with the Institut de Microbiologie de l'U. de Montreal, in Montreal, Que., Canada.

Professor Robert Dorfman has been appointed as Associate Professor in the Department of Economics, Harvard University.

Benjamin Epstein has been promoted to a full professorship at Wayne University. During the 1955–56 academic year he is on Sabbatical leave at Stanford University.

W. T. Federer returned to Cornell University in September, 1955, after one year, from Honolulu, Hawaii, where he was head of the Department of Experimental Statistics, Experiment Station, Hawaiian Sugar Planters' Association and a consultant in Statistics for the Pineapple Research Institute.

Dr. Rudolf J. Freund, who received his Ph.D. degree in statistics at North Carolina, has joined the staff of the Department of Statistics as Associate Professor at the Virginia Polytechnic Institute.

Research Professor Bernard Friedman of New York University is on leave of absence and has been appointed to a visiting professorship at the University of California, Berkeley.

After receiving a Ph.D. in Statistics from the University of North Carolina June 1955, Seymour Geisser was at the Statistical Engineering Laboratory, N.B.S. Washington, D. C., before accepting a commission with the U. S. Public Health Service four months later. He is now with the N.I.H. at Bethesda, Maryland.

Leo A. Goodman, formerly Associate Professor, has been promoted to Professor of Statistics and Sociology at the University of Chicago.

Earl L. Green has returned to his position as professor of genetics at Ohio State University after being on leave of absence for two years while serving as Geneticist, Division of Biology and Medicine, U. S. Atomic Energy Commission.

Bernard Harris is on leave from his position as mathematician, Defense Department, and Instructor in Statistics, George Washington University, and will be at the Stanford University Applied Mathematics and Statistcs Laboratory for one year.

Paul G. Homeyer has been in Mexico for most of the fall quarter of 1955 as agricultural statistician with the Food and Agriculture Organization, United Nations. He returns to the Statistical Laboratory, Iowa State College, in December.

Robert H. Hoskins of the John Hancock Mutual Life Insurance Company recently received an appointment as Assistant Group Actuary.

Dr. Stanley Isaacson has resigned his position as Assistant Professor of Statistics at Iowa State College in order to accept a position as a Statistician with the Semiconductor Department of Westinghouse Electric Corporation.

Dr. Koichi Ito, formerly of Institute of Statistics, University of North Carolina, returned to Japan in August, 1955, and is now Lecturer in Statistics at Nanzan University, Nagoya, Japan.

Ralph B. Johnson resigned from Clemson College and accepted a position as Assistant Professor of Mathematics at Catawba College, Salisbury, North Carolina.

After receiving his masters degree from Iowa State College August 1955, Lawrence F. Jones accepted a position in the Statistical Control Section of the Semiconductor Department of the Westinghouse Electric Corporation.

Dr. Tosio Kitagawa joined the Statistical Laboratory of Iowa State College as visiting professor for the fall quarter of 1955. He has been on leave from the position of professor of theory of probability and mathematical statistics at Kyusyu University, Fukuoka, Japan, since last summer.

Dr. Morris Krauss has begun postdoctoral research at the National Bureau

of Standards. He is one of seven young graduate scientists to be selected for the Postdoctoral Research Associateship program sponsored by the National Academy of Sciences-National Research Council and the Bureau.

Boyd Ladd has been appointed research manager of Southwest Research Institute's department of industrial economics.

Ferdinand Lemus, formerly a graduate assistant at the Statistical Laboratory, Iowa State College, has accepted a position with the Experimental Design and Statistical Analysis Group of Westinghouse Electric Corporation, Pittsburgh, Pennsylvania.

Craig A. Magwire has joined the staff of the Department of Mathematics and Mechanics at the U. S. Naval Postgraduate School, Monterey, California, as Associate Professor of Mathematics.

Allen L. Mayerson has completed his year as a Fullbright scholar at the Sorbonne, and has returned to his position as Principal Actuary of the N. Y. State Insurance Dept. While abroad he was elected a member of the Swiss Actuarial Society and of the French Institute of Actuaries, and did some research into European actuarial practices, in addition to his mathematical studies at the Sorbonne.

Franklin S. McFeely has completed the work for the Ph.D. in Statistics at Virginia Polytechnic Institute and is now on the staff of the University of Colorado School of Medicine.

Mrs. Mary G. Natrella has rejoined the Applied Mathematics Division of the National Bureau of Standards where she will serve on the staff of the Statistical Engineering Laboratory.

Joseph A. Navarro received his Ph.D. in mathematical statistics from Purdue University in August 1955 and is at present employed at the General Electric Advanced Electronics Center in Ithaca, New York.

John Neter, formerly of Syracuse University, has accepted an appointment as Associate Professor in the School of Business Administration, University of Minnesota, Minneapolis, Minnesota.

Dr. Richard E. Nettleton has begun postdoctoral research at the National Bureau of Standards. He is one of seven young graduate scientists to be selected for the Postdoctoral Research Associateship program sponsored by the National Academy of Sciences-National Research Council and the Bureau.

Bernard Ostle has been promoted to Professor of Mathematics and Statistics at Montana State College.

Gilbert I. Paul, formerly a student at N. C. State College, Raleigh, N. C. , is now teaching courses in statistics and in statistical genetics at McGill University, Montreal, Canada.

Dr. Dayle D. Rippe has accepted a position as Operations Research Analyst in the General Office of General Mills, Inc., Minneapolis. During the three years prior to this he was an Operations Analyst with the Strategic Air Command, Omaha, Nebraska.

Jagdish Sharon Rustagi has completed his work for Ph.D. degree in Statistics

at Stanford University, California, and has a teaching job in the Department of Mathematics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, for the academic year, 1955–56.

Richard H. Shaw resigned from the U. S. Naval Ordnance Plant in Indianapolis, to take a position with General Dry Batteries, Inc., in Cleveland.

Jack Silber of Roosevelt University spent the summer of 1955 as Consultant to the Operations Analysis Office at the U. S. Air Force Missile Test Center, Patrick AFB, Florida.

Upon completion of his Ph.D. in the Dept. of Mathematics at Purdue University, John A. Tischendorf received an appointment with the Commissioned Corps of the U. S. Public Health Service. He is now stationed at Boston, Massachusetts, serving as statistician on a study conducted by the Chronic Disease Program.

Donald R. Truax has been appointed Research Fellow in Mathematics at the California Institute of Technology after receiving his Ph.D. in statistics at Stanford University.

Howard G. Tucker received his Ph.D. at the University of California in June, 1955, and is now Assistant Professor of Mathematics at the University of Oregon.

Mr. Robert S. Walleigh has rejoined the staff of the National Bureau of Standards as Assistant Director for Administration. In this position, he will serve as the Director's principal staff advisor on management matters, and supervise the operation of the administrative divisions that support the Bureau's technical program.

Dr. R. Lowell Wine, who received his Ph.D. degree in statistics in June 1955 from the Virginia Polytechnic Institute, has joined the staff there in the Department of Statistics as Associate Professor.

Hans-Joachim Zindler has been appointed as consultant for Mathematical Statistics in the "Statistisches Bundesant" Abteilung VIII, in Wiesbaden, Germany.

New Members

The following persons have been elected to membership in the Institute

August 6, 1955 to November 9, 1955

Anderson, Allan G., Ph.D. (University of Michigan), Mathematician, Jones and Laughlin Steel Corporation, Research Center, 900 Agnew Road, Pittsburgh 30, Pennsylvania.

Baldwin, Roger R., M. A. (Columbia University), Graduate Student, Princeton University, Mathematics Department, Fine Hall, Princeton, New Jersey, *282 W. 11th Street, New York 14, New York.*

Banerjee, D. P., D.Sc. (Dacca University) Department Head of Mathematics, Meerut College, Meerut U.P., India.

Bessler, Stuart Alan, Bachelor of Industrial Engineering and Bachelor of Business Administration (University of Minnesota), Graduate student and instructor, Depart-

ment of Economics, School of Business Administration, University of Minnesota, Minneapolis 14, Minnesota, *3922 Basswood Road, Minneapolis 16, Minnesota.*

Bissinger, Barnard Hinkle, Ph.D. (Cornell University), Chairman Mathematics Dept., Lebanon Valley College, Annville, Pennsylvania.

Brenna, Leroy S., M.S., (Kansas State College), Industrial Engineer, Eastman Kodak Company, 343 State Street, Rochester, 4, New York, *Strathmore Circle, Bldg. 6, Apartment 3, Rochester 9, New York.*

Busch, Sister Mary Constance, C.S.A., B.S. in Ed. (Marian College), Math. Instructor, Marian College, 30 East Division Street, Fond du Lac, Wisconsin.

Dempster, Arthur P., M.A. (Univ. of Toronto), Student, Princeton Univ., Dept. of Mathematics, Fine Hall, Box 708, Princeton, New Jersey.

Drenick, Rudolf F., Ph.D. (Univ. of Vienna), Manager Analytical Group, Radio Corporation of America, Bldg. 10–8, Camden 2, New Jersey.

Ferris, George Emery, M.A. (Columbia Univ.), Graduate Assistant, Cornell University, Ithaca, New York, *114 Eddy Street, Ithaca, New York.*

Feyerherm, Arlin M., Ph.D. (Iowa State College), Assistant Professor of Mathematics, Kansas State College, Manhattan, Kansas.

Fiske, Edwards N., B.S. (Roanoke College), Graduate Student, Virginia Polytechnic Institute, Blacksburg, Virginia, *Box 3611, Virginia Tech. Station V, Blacksburg, Virginia*

Gregoire, Jean (Mr.), B.Sc. (Laval University), Graduate Student and assistant, University of Manitoba, Winnipeg, Manitoba, Canada, *389 6th Avenue, Grand Mere, Que. Canada.*

Hadley, Hershel N., B.A. (Whitman College), Head, Statistical Division, Inspection and Test Department, Naval Powder Factory, Indian Head, Maryland, *5706 Fortieth Avenue, Hyattsville, Maryland.*

Helms, Lester L., M.S. (Purdue Univ.), Mathematician, Operations Research Group, Convair, Pomona, California.

Hermann, Philip, M.S. (Case Institute of Technology), Supervisor Applied Mathematics, Jones and Laughlin Steel Corporation, No. 3 Gateway Center, Pittsburgh, Pa.

Kassler, Raymond, B.A. (Brooklyn College), Research Engineer, Boland and Boyce, Inc., Belleville 9, New Jersey, *P.O. Box 441, Camden 1, New Jersey.*

Keats, Mortimer B., M.A. (George Washington University), Head, Statistics and Analysis Division, Quality Surveillance Department, U. S. Naval Powder Factory, Indian Head, Maryland, *2354 Skyland Place, S.E., Washington 20, D. C.*

Miller, Rupert Griel, Jr., A. B. (Princeton), Student, Stanford University, Statistical Lab., Sequoia Hall, Stanford, California.

Ogawa, Junjiro, Ph.D. (Osaka University), Asst. Prof., Department of Mathematics, Faculty of Science, Osaka University, Japan, *% The Dept. of Statistics, Univ. of North Carolina, Chapel Hill, N. C.*

Remmenga, Elmer E., Ph.D. (Purdue Univ.) Asst. Prof., Department of Mathematics, Colorado A. and M., Fort Collins, Colorado., *1821 Crestmore Place, Fort Collins, Colorado.*

Roberts, Howard R., B.Sc. (George Washington Univ.) Graduate Teaching Assistant, George Washington University, Department of Statistics, Washington 6, D.C., *2022 G Street, N.W. Washington 6, D. C.*

Romani Miquel, José, Dipl. en Estadistica (Madrid Univ.), Colaborador, Instituto de Investigaciones Estadisticas, Serrano 123, Madrid, Spain, *Harzenbusch 5, Madrid, Spain.*

Ryan, John M., Ph.D. (University of North Carolina), Mathematical Economist, United Gas Corporation, P. O. Box 1407, Shreveport, Louisiana.

Sakaguchi, Minoru, (Tokyo Institute of Technology), Lecturer, Department of Mathematics, University of Electro-communication, Kojima-cho 14, Chohu, Tokyo, Japan.

Schmid, Paul, Diplom der ETH (Eidg. Anstalt fur das forstliche Versuchswesen), Research

Assistant ETH, Dipl. Math. ETH, Mathematician, Eidg. Anstalt fur das forstliche Versuchswesen, Tannenstr. 11, Zurich 6, Switzerland.

Siegel, Sidney, Ph.D. (Stanford Univ.) Asst. Prof., Department of Psychology, Pennsylvania State University, University Park, Pennsylvania.

Shtulman, Sidney, BBA (College of City of New York), Analytical Statistician, The Theory and Analysis Division of the Computation and Ballistics Department Federal Civil Service, Dahlgren, Virginia, *Naval Proving Ground, Dahlgren Va.*

Siller, Harry, Ph.D. (New York University), Asst. Professor, Department of Mathematics, Hofstra College, Hampstead, New York, *1139 Beach Channel Drive, Far Rockaway, New York.*

Snell, James Laurie, Ph.D. (Univ. of Illinois), Asst. Prof., Department of Mathematics, Dartmouth College, Hanover, N. H., *23 S. Park St., Hanover, N. H.*

Stenwick, Fern Caroline, Mathematician, David Taylor Model Basin, Carderock, Maryland, *1039 Columbia Drive, Bucknell Manor, Alexandria, Virginia.*

Taylor, Robert J., M. S. (Virginia Polytechnic Institute), Mathematician, Naval Research Laboratory, Washington 25, D. C., *4761-B S. Capitol Terrace, S. W., Washington 24, D. C.*

Thomas, Raymond E., B. A. (George Washington University), Student, Graduate Assistant, George Washington University, Washington 6, D. C., *4249 Hildreth St., S. E., Washington 19, D. C.*

Trinkl, Frank N., A. M. (University of Michigan), Student, Graduate Assistant, Stanford University, Stanford, California, *213-14 Stanford Village, Stanford, Calif*

Turner, Nura Dorothea, M.S. (State Univ. of Iowa), Asst. Prof., Department of Mathematics, New York State College for Teachers, Albany, New York.

Wolinez, G., M.Sc. (Hebrew University), Research Worker, Government of Israel, Hahiriah, Tel-Aviv, Israel, *Dizzengoff 204, Tel-Aviv, Israel.*

Cooperative Graduate Summer Sessions in Statistics

The University of Florida, North Carolina State College, Virginia Polytechnic Institute and the Southern Regional Education Board are jointly sponsoring a series of cooperative summer sessions in statistics.

The third of these summer sessions will be held at North Carolina State College, June 11–July 20, 1956. A session is scheduled to be held at Virginia Polytechnic Institute in 1957 and at the University of Florida in 1958. Each summer session lasts six weeks and each course carries approximately three semester hours of graduate credit.

The 1956 session will be held jointly with the Institute in Quantitative Research Methods in Agricultural Economics, sponsored by the Social Science Research Council. Several statistics courses will be oriented towards economic applications.

The combined faculty for the 1956 summer session and Institute at North Carolina State College will include: Professor R. L. Anderson, North Carolina State College; Professor Gertrude M. Cox, North Carolina State College; Professor David B. Duncan, University of Florida; Professor Alva L. Finkner, North Carolina State College; Dr. Arnold H. E. Grandage, North Carolina State College; Professor Robert J. Hader, North Carolina State College; Assistant Professor Cleon Harrell, North Carolina State College; Professor Earl O. Heady,

Iowa State College; Professor Clifford G. Hildreth, Michigan State University; Professor Jack Levine, North Carolina State College; Professor Robert J. Monroe, North Carolina State College; and Assistant Professor Walter L. Smith, University of North Carolina.

Courses to be offered this summer are: Statistical Methods I, Statistical Methods II (Design of Experiments), Statistical Theory I (Probability and Parent Distribution), Statistical Theory II (Sampling Distributions and Inference), Sample Survey Designs, Advanced Analysis II, Advanced Calculus for Statistics, Stochastic Processes, Econometric Methods and Linear Programming. Lectures on Linear Equations (Matrix Algebra) and Production Functions will be given in the Institute program.

Inquiries should be addressed to:

Professor J. A. Rigney
Department of Experimental Statistics
North Carolina State College
Raleigh, North Carolina

Summer Sessions at Berkeley, California

The 1956 summer program in the Department of Statistics of the University of California, Berkeley, California, will consist of two sessions: June 18 to July 28 and July 30 to September 8. The faculty of the summer sessions will include Professor D. R. Cox of the University of North Carolina, Professor Grace E. Bates of Mount Holyoke College, and Professor David Blackwell and Mr. T. S. Ferguson of the Department of Statistics of the University of California.

The program includes two of the usual undergraduate courses in each session, adapted primarily to meet the needs of students transferring from other centers who would like to undertake advanced study at the University of California during the regular academic year. Also a graduate seminar will be conducted by Professor Blackwell. This seminar will allow for individual consultation for students working toward higher degrees.

Fellowship in Experimental Statistics

The Department of Industrial and Engineering Administration of the Sibley School of Mechanical Engineering at Cornell University announces a graduate fellowship in the area of experimental statistics, sponsored by the Standard Oil Company of Ohio. This fellowship carries a stipend of $2,250 for the academic year and is to be used to support work toward either the M.S. or Ph.D. degree.

Applicants are expected to have suitable undergraduate preparation in some branch of engineering or in the physical sciences, but undergraduate training in statistics is not a prerequisite. The holder of the fellowship will be expected to

take courses in both applied and mathematical statistics selected from among the offerings of the members of the Cornell Statistics Center.

Interested persons who desire additional information concerning this fellowship may write to Professor Robert E. Bechhofer at the above address. Application forms for this fellowship and for admission to the graduate school may be obtained from the Graduate School, 125 Edmund Ezra Day Hall, Cornell University, Ithaca, New York, and should be filed not later than February 17, 1956, for April 1 award. Late applications will be considered only if an award has not been made by April 1.

REPORT OF THE NEW YORK MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The sixty-eighth meeting of the Institute of Mathematical Statistics and the eighteenth annual meeting was held at the Hotel Biltmore, New York City, on December 27–30, 1955, in conjunction with the national annual meeting of the American Statistical Association and the Biometric Society (ENAR). A number of the sessions were joint with these two organizations. A Special Invited Paper entitled *Stochastic Approximation* was presented by Professor Cyrus Derman of Columbia University. The Rietz Lecture, entitled *Probability in Statistics*, was given by Professor William Feller of Princeton University. The following members of the Institute attended:

Forman S. Acton, Beatrice Aitchison, William Robert Allen, John E. Alman, Richard L. Anderson, Theodore W. Anderson, Fred C. Andrews, Harvey James Arnold, Herbert E. Arnold, Kenneth J. Arnold, Kenneth J. Arrow, Max Astrachan, R. R. Bahadur, J. C. Bain, Roger Raushenbush Baldwin, James B. Bartoo, Grace E. Bates, Geoffrey Beall, Helen P. Beard, Robert E. Bechhofer, Gordon H. Beckhart, Agnes Berger, Abraham J. Berman, A. T. Bharucha-Reid, Arnold M. Binder, Richard S. Bingham, Allan Birnbaum, Z. William Birnbaum, David Blackwell, Archie Blake, Chester I. Bliss, Aaron Block, Julius R. Blum, Isadore Blumen, Robert M. Blumenthal, Paul Boschan, Raj. C. Bose, Albert H. Bowker, Ralph Allan Bradley, A. E. Brandt, Irwin D. J. Bross, Benjamin Buchbinder, Robert Wilbur Burgess, Paul J. Burke, Lyle D. Calvin, Mavis B. Carroll, Maria Castellani, Douglas G. Chapman, A. Charnes, Herman Chernoff, John T. Chu, Ira H. Cisin, Willard H. Clatworthy, Paul C. Clifford, William G. Cochran, Theodore Colton, William Stokes Connor, Jerome Cornfield, Louis J. Cote, David R. Cox, Edwin Lory Cox, Gertrude M. Cox, Paul Charles Cox, Cecil C. Craig, Jean B. Crockett, Edwin Louis Crow, Lee Crump, Phelps P. Crump, Paz B. Culabutan, Joseph Francis Daly, Cuthbert Daniel, Reed B. Dawson, Jr., Besse B. Day, Ralph A. DeMarr, Francis R. Del Priore, Lucile Derrick, James L. Dolby, Thomas Donnelly, Robert Dorfman, Joseph Abraham Dresner, Acheson J. Duncan, David B. Duncan, Mary T. Dunleavy, Charles

William Dunnett, David Durand, Arthur M. Dutton, Meyer Dwass, A. Ross Eckler, George L. Edgett, Sylvain Ehrenfeld, Harvey Eisenberg, Harry Eisenpress, Henry Ellner, Daniel R. Embody, Benjamin Epstein, Robert Marvin Exselsen, Walter T. Federer, William Feller, Thomas S. Ferguson, Robert Ferber, George Emery Ferris, John W. Fertig, Edwards Ned Fiske, John C. Flanagan, Lester R. Frankel, David Frazier, Spencer M. Free, Jr., Harold A. Freeman, John E. Freund, Henry D. Friedman, Milton Friedman, Fred Frishman, L. A. Gardner, Jr., Norman Robert Garner, Murray A. Geisler, Seymour Geisser, J. Lincoln Gerende, Dorothy M. Gilford, Leon Gilford, Ruth L. Gold, Leo A. Goodman, Nathaniel R. Goodman, Mina Haskind Gourary, Franklin A. Graybill, Bernard G. Greenberg, Samuel W. Greenhouse, Joseph A. Greenwood, Geoffrey Gregory, Thomas N. E. Greville, Harold Gulliksen, Emil J. Gumbel, Lee Gunlogson, Paul Gunther, Margaret Gurney, Robert John Hader, John S. Hagan, Keet W. Halbert, William Jackson Hall, Max Halperin, E. Cuyler Hammond, James F. Hannan, Morris H. Hansen, Gordon M. Harrington, Theodore E. Harris, Boyd Harshbarger, Herman O. Hartley, David G. Hays, William Carleton Healy, Jr., Ernest Earl Heinbach, Paul Heit, Leon H. Herbach, G. Ronald Herd, Philip Hermann, Irene Hess, Werner Hochwald, William Hodgkinson, Jr., Wassily Hoeffding, Paul G. Homeyer, William C. Hood, Robert Hooke, John W. Hopkins, Daniel G. Horvitz, Harold Hotelling, Earl E. Houseman, William Gerow Howe, J. Stuart Hunter, Frederick Vincent Hurst, Jr., Paul E. Irick, Nathan Jaspen, Raymond J. Jessen, Milton Vernon Johns, Jr. Howard L. Jones, Hyman B. Kaitz, Edward L. Kaplan, A. E. Karp, Marvin Aaron Kastenbaum, Leo Katz, Lester S. Kellogg, Oscar Kempthorne, Robert W. Kennard, George H. Kennedy, Jack C. Kiefer, Allyn W. Kimball, Bradford F. Kimball, Leslie Kish, Tosio Kitagawa, Tjalling C. Koopmans, Richard L. Kozelka, Charles H. Kraft, Evelyn Lucille Kramer, William Henry Kruskal, Roy R. Kuebler, Jr., Thomas E. Kurtz, Jack Laderman, Helen Humes Lamale (Mrs.), Andre G. Laurent, Fred Charles Leone, Howard Levene, Edward Abraham Lew, Everett Vernon Lewis, Gerald J. Lieberman, Gilbert Lieberman, Jacob E. Lieberman, Julius Lieblein, Rensis Likert, Richard F. Link, Benjamin Lipstein, Sebastian B. Littauer, Eugene Lukacs, Robert James Lundegard, George F. Lunger, Philip J. McCarthy, Brockway McMillan, Robert G. McMillan, Gertrude A. McQuaid, Ralph L. Madison, Benjamin Malzberg, John Mandel, Joseph Mandelson, Herbert Marshall, Frank J. Massey, Paul Meier, W. J. Merrill, Herbert A. Meyer, Paul L. Meyer, Paul D. Minton, Irwin Miller, Robert Mirsky, Azizali Farrukh Mohammed, Edward C. Milina, A. M. Mood, Milton Morrison, Norman Morse, Joseph Edward Morton, Lincoln E. Moses, Jack Moshman, Frederick Mosteller, Shu-Teh Chen Moy, Merv E. Muller, Ray B. Murphy, Luis F. Nanni, Tadepalli Vankata Narayana, Mary G. Natrella, Joseph Anthony Navarro, M. R. Neifeld, A. Carl Nelson, Jr., Morton J. Netzorg, Robert Joseph Nichol, George E. Nicholson, Jr., Wesley L. Nicholson, Nicholas Nikitich, Harold Nisselson, Lionel MacLean Noel, Gottfried E. Noether, Monroe L. Norden, Nilan Norris, Horace W. Norton, James Norton, Jr., Junjiro Ogawa,

Carl Reading Ohman, Edwin G. Olds, Ingram Olkin, Paul S. Olmstead, A. L. O'Toole, Donald B. Owen, William R. Pabst, Jr., Emanuel Parzen, Robert E, Patton, Edward Burton Perrin, John K. Perrin, Eusche W. Pike, John W. Pratt, Lila Knudsen Randolph, Bayard Rankin, Stanley Reiter, George J. Resnikoff, Joseph S. Rhodes, Wyman Richardson, Donald L. Richter, Paul R. Rider, William L. Roach, Jr., Herbert E. Robbins, Helen Murray Roberts, Selby Lemley Robinson, Douglas S. Robson, Robert Roeloffs, Albert C. Rohloff, Charles F. Roos, John H. Roseboom, David Rosenblatt, Harry M. Rosenblatt, Joan Raup Rosenblatt, Judah Isser Rosenblatt, Murray Rosenblatt, Willard C. Ross, S. N. Roy, Jagdish S. Rustagi, John Morris Ryan, David Sachs, Jerome Sacks, Daniel E. Sands, F. E. Satterthwaite, L. J. Savage, Edward Sax, Henry Scheffé, Marvin A. Schneiderman, Elizabeth L. Scott, Hilary Seal, Oliver Abbott Shaw, Richard H. Shaw, Sidney Shtulman, Elizabeth Anna Shuhany, Sidney Siegel, Walter R. Simmons, Monroe Gilbert Sirken, Rosedith Sitgreaves, Morris Skibinsky, Hugh Fairfield Smith, John H. Smith, Walter Laws Smith, Jean F. Smolak, James L. Snell, Milton Sobel, Herbert Solomon, Paul N. Somerville, Frederick A. Sorensen, D. E. South, Mortimer Spiegelman, B. Ralph Stauber, George Powell Steck, Robert G. D. Steel, Arthur Stein, Frederick F. Stephan, Leo Eugene Storm, Fred W. Strodtbeck, Hale C. Sweeny, Zenon Szatrowski, Francis B. Taylor, Henry Teicher, Dan Teichroew, Milton E. Terry, Gerhard Tintner, Hebert C. S. Thom, Donovan J. Thompson, William Rae Thompson, Leo J. Tick, Mary Newton Torrey, Chia Kwei Tsao, Albert William Tucker, John W. Tukey, Charles R. M. Tuttle, M. C. K. Tweedie, Jose Vergara, D. F. Votaw, Jr., Helen M. Walker, David L. Wallace, W. Allen Wallis, Sidney Weiner, Harry Weingarten, Eleanor S. Weiss, Irving Weiss, Lionel Weiss, Oscar Wesler, Phillips Whidden, John S. White, Alfred G. Whitney, Frank Wilcoxon, Martin B. Wilk, R. Lowell Wine, B. J. Winer, Gerald Winston, William Wolman, Max A. Woodbury, Charles Ashley Wright, John William Youden, Samuel Zahl, Royal Keith Zeigler, Marvin Zelen.

The program follows:

TUESDAY, DECEMBER 27, 1955

9:00 a.m. Contributed Papers I

Chairman: Rosedith Sitgreaves, Teachers College, Columbia University

Papers:

1. *The Midrange of a Sample as an Estimator of the Population Midrange*, Paul R. Rider, Wright-Patterson Air Force Base.
2. *Distribution of the Product of Maximum Values in Samples from a Rectangular Population*, Paul R. Rider, Wright-Patterson Air Force Base, (By title).
3. *A Note on Non-recurrent Random Walks*, Cyrus Derman, Columbia University, (By title).
4. *Statistical Spectral Analysis, I: Consistent Asymptotically Normal Estimates of the Covariance Function and Spectral Averages*, Emanuel Parzen, Columbia University, (By title).
5. *Statistical Spectral Analysis, II: Asymptotic Mean Square Error of a Class*

of Estimates of the Spectral Density, EMANUEL PARZEN, Columbia University.

6. *A Central Limit Theorem for Multilinear Stochastic Processes*, EMANUEL PARZEN, Columbia University, (By title).
7. *An Extension of Cramer's Theorem 20.6 to Random Functions with Values in a Metric Space*, EMANUEL PARZEN, Columbia University, (By title).
8. *Orthogonality and Fractional Replication of Factorial Experiments*, ALLAN BIRNBAUM, Columbia University.
9. *On the Second Sample Size Function of a Bayes Two-Stage Test for the Mean*, MORRIS SKIBINSKY, Purdue University.
10. *A New Estimation Procedure for a Linear Combination of Exponentials (Preliminary Report)*, RICHARD G. CORNELL, Oak Ridge National Laboratory and Virginia Polytechnic Institute.
11. *A Note on Weighted Randomization*, D. R. COX, University of North Carolina.
12. *On the Analysis of Incomplete Block Designs*, MARVIN ZELEN, National Bureau of Standards.
13. *A Remark on Wald's Paper, "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups,"* JUNJIRO OGAWA, University of North Carolina.
14. *Consistency and Optimum Properties of Some Two-Sample Tests*, JULIUS R. BLUM, Indiana University, and LIONEL WEISS, University of Virginia.

11:00 a.m. Probability and Statistics in Genetics. With Biometric Society (ENAR)

Chairman: OSCAR KEMPTHORNE, Iowa State College

Papers:
1. *Some Problems of Stochastic Processes in Genetics*, M. KIMURA, University of Wisconsin.
2. *Estimation of Parameters in Genetic Models*, HOWARD LEVENE, Columbia University.
3. *Sequential Tests for Detection of Linkage in Man*, NEWTON E. MORTON, University of Wisconsin.

2:00 p.m. Contributions of M. A. Girshick to Mathematical Statistics. With American Statistical Association

Chairman: HENRY SCHEFFÉ, University of California

Papers:
1. *Multivariate Analysis*, HAROLD HOTELLING, University of North Carolina.
2. *Sequential Analysis*, EDWARD PAULSON, Queens College, New York City.
3. *Decision Theory*, DAVID BLACKWELL, University of California.

4:00 p.m. Extreme Value Theory

Chairman: SEBASTIAN B. LITTAUER, Columbia University

Papers:
1. *Statistical Theory of Fatigue Failure and Breaking Strength*, E. J. GUMBEL, Columbia University.
2. *On the Problem of Forecasting Extreme Values from a Curve Fitted to the Type I Extreme Value Distribution*, BRADFORD F. KIMBALL, New York State Public Service Commission.
3. *Developments in the Application of Extreme Value Theory*, JULIUS LIEBLEIN, National Bureau of Standards.

Discussion: BENJAMIN EPSTEIN, Wayne University and Stanford University, C. H. S. THOM, Advisory Committee on Weather Control

4:00 p.m. Components of Variance. With the American Statistical Association

Chairman: C. W. Dunnett, American Cyanamid Company

Papers: 1. *Components of Variance, Finite Populations, and Statistical Inference*, H. F. Smith, North Carolina State College.
2. *Non-additivity in a Latin Square Design*, M. B. Wilks nd O. Kempthorne, Iowa State College.

Discussion: John Tukey, Princeton University

6:00 p.m. 1955 Council Meeting

Chairman: Henry Scheffé, President

WEDNESDAY, DECEMBER 28, 1955

9:00 a.m. Contributed Papers II

Chairman: M. Vernon Johns, Columbia University

Papers:
1. *Remarks on Characteristic Functions*, Eugene Lukacs, Catholic University of America and Office of Naval Research.
2. *The Limiting Distribution of Serial Correlation Coefficient in the Explosive Case*, John S. White, University of Manitoba.
3. *The Distribution of the Ratio of Two Measures of Normal Dispersion*, H. O. Hartley, Iowa State College.
4. *Estimating a Linear Functional Relation*, H. Fairfield Smith, North Carolina State College.
5. *Asymptotic Distribution of Roots of Certain Determinantal Equations*, R. Gnanadesikan, University of North Carolina, (By title).
6. *Investigation of the Possibility of Using Likelihood Ratio Tests of Certain Multivariate Hypotheses for Obtaining Confidence Bounds*, R. Gnanadesikan, University of North Carolina, (By title).
7. *Asymptotic Efficiencies of a Nonparametric Life Test for Investigating Smaller Percentiles of a Gamma Distribution*, John E. Walsh, Lockheed Aircraft Corporation, (By title).
8. *A Test of Judge Concordance for Paired Comparison Designs (Preliminary Report)*, J. W. Wilkinson, University of North Carolina, (By title).
9. *On the Efficiency of Certain Classes of Tests Based on the U-Statistics*, Joan Raup Rosenblatt, National Bureau of Standards.
10. *The Dynamic Statistical Decision Problem when the Component Problem Involves a Finite Number, m, of Distributions*, James F. Hannan, Michigan State University.
11. *On Certain Systems of Experiments as Interdependent Stochastic Processes (Preliminary Report)*, David Rosenblatt, American University, (By title).
12. *A Spherically-Symmetric Order Statistic r, (Preliminary Report)*, Brian Gluss and Fred L. Strodtbeck, University of Chicago.

11:00 a.m. Statistical Studies of Accident Proneness and Contagion of Accidents. With the Biometric Society (ENAR)

Chairman: M. Vernon Johns, Columbia University

Papers: 1. *General Review of Recent Results on Accident Proneness and Contagion*, J. Neyman, University of California.

2. *Asymptotic Tests and Power of Tests of Certain Hypotheses Regarding Contagion in Accidents*, CHARLES H. KRAFT, University of California.
3. *A Limit Theorem on Conditional Distributions Related to Studies of Accident Proneness*, G. P. STECK, Sandia Base.

2:00 p.m. Rietz Lecture

Chairman: J. NEYMAN, University of California

Paper: *Probability in Statistics*, WILLIAM FELLER, Princeton University.

4:00 p.m. Acceptance Sampling Plans

Chairman: GERALD J. LIEBERMAN, Stanford University

Papers:
1. *Some Continuous Sampling Plans for Complex Items*, GEORGE RESNIKOFF, Stanford University.
2. *An Economic Approach to the Choice of Continuous Sampling Plans*, GEOFFREY GREGORY, Stanford University.
3. *On the Construction of Optimum Double Sampling Plans*, ALLAN BIRNBAUM, Columbia University.

6:00 p.m. Business Meeting

Chairman: HENRY SCHEFFÉ, President

8:30 p.m. Council Meeting

Chairman: DAVID BLACKWELL, President

THURSDAY, DECEMBER 29, 1955

9:00 a.m. A Training Program Leading to Contributions in Experimental Design

Chairman: CHURCHILL EISENHART, National Bureau of Standards

Papers:
1. *Who Makes Designs*, W. J. YOUDEN, National Bureau of Standards.
2. *The Construction of Fractional Factoral Designs for the 2^n Series*, F. L. MILLER, JR., Purdue University and National Bureau of Standards.
3. *The Use of the 2^n Fractional Factorial Designs for Factors at 4 Levels*, H. PETTIGREW, George Washington University and National Bureau of Standards.
4. *Some Combinatorial Relationships Arising in the Dualization of Incomplete Block Designs*, R. BURTON, National Bureau of Standards.

Discussion: GERTRUDE M. COX, North Carolina State College.

11:00 a.m. Statistics in Medical Experimentation. With the Biometric Society (ENAR)

Chairman: LINCOLN E. MOSES, Stanford University

Papers:
1. *Elimination of Selection Bias in Medical Experimentation*, DAVID BLACKWELL and J. L. HODGES, Jr., University of California.
2. *Estimation of Bacterial Densities*, THOMAS S. FERGUSON, University of California.
3. *On Comparing Survival Rates*, AGNES BERGER, School of Public Health, Columbia University.

2:00 p.m. Special Invited Paper

Chairman: HERMAN CHERNOFF, Stanford University

Paper: *Stochastic Approximation*, CYRUS DERMAN, Columbia University.

4:00 p.m. Recognized Needs for Mathematical Tables among Statisticians

Chairman: ALBERT BOWKER, Stanford University and Columbia University

Discussion Session

4:00 p.m. New Developments in Experimental Social Science. With the Social Statistics Section of the American Statistical Association

Chairman: FRED L. STRODTBECK, University of Chicago

Papers:
1. *Monte Carlo Methods in an Experimental Test of an Interaction Model*, DAVID G. HAYES, The RAND Corporation.
2. *Some Theoretical Problems of Lexico-Statistics*, MORRIS SWADESH, Denver, Colorado. (Presented by Joseph H. Greenberg, Columbia University.)
3. *Mathematical Models for the Empirical Study of Decision-Making*, PATRICK SUPPES, Stanford University.

Discussion: HERBERT SOLOMON, Columbia University
LEONARD J. SAVAGE, University of Chicago
DAVID R. COX, University of Cambridge.
JOSEPH H. GREENBERG, Columbia University.

9:30 p.m. Informal Party. With American Statistical Association

FRIDAY, DECEMBER 30, 1955

9:00 a.m. Contributed Papers III

Chairman: S. N. ROY, University of North Carolina

Papers:
1. *Generalized Normalization Polynomials*, D. TEICHROEW, University of California at Los Angeles.
2. *Tables for Computing Bivariate Normal Probabilities*, DONALD B. OWEN, Sandia Corporation.
3. *Bounds and Approximations for Constants Used in Quality Control*, J. T. CHU, University of North Carolina and Case Institute of Technology.
4. *Four Streams of Traffic Converging on a Cross-Road*, BRIAN GLUSS, University of Chicago, introduced by FRED L. STRODTBECK, University of Chicago, (By title).
5. *Markov Processes Arising in Learning Models*, JOHN G. KEMENY, and J. L. SNELL, Dartmouth College.
6. *On a Decision Rule for Selecting a Group Containing the Population with the Largest Mean (Preliminary Report)*, R. C. BOSE and S. S. GUPTA, University of North Carolina, (By title).
7. *Recurrent Values of Sums of Independent Random Variables (Preliminary Report)*, LOUIS J. COTE and HENRY TEICHER, Purdue University.
8. *A Problem Involving the Distribution of Shadows (Preliminary Report)*, HERMAN CHERNOFF, Stanford University, and JOSEPH F. DALY, National Bureau of the Census.
9. *Note on Two-Stage Test Procedures*, S. G. GHURYE, Lucknow University, (By title).
10. *Some Properties of Generalized Sequential Probability Ratio Tests*, J. KEIFER, Cornell University, and LIONEL WEISS, University of Virginia.
11. *Sequential Decision Problems for a Class of Stochastic Processes. Testing Hypotheses (Preliminary Report)*, A. T. BHARUCHA-REID, University of California.

12. *Note on a Markov Chain with Matrix States and Some Applications*, A. T. Bharucha-Reid, University of California, (By title).
13. *On the Comparison of Two Stochastic Epidemics*, A. T. Bharucha-Reid, University of California, (By title).
14. *A Sequential Multiple Decision Procedure for Ranking Means of Normal Populations with a Common Unknown Variance (Preliminary Report)*, R. E. Bechhofer, Cornell University, and M. Sobel, Bell Telephone Laboratories, (By title).
15. *A Scale-Invariant Sequential Multiple Decision Procedure for Ranking Variances of Normal Populations (Preliminary Report)*, R. E. Bechhofer, Cornell University, and M. Sobel, Bell Telephone Laboratories, (By title).
16. *Exact Probabilities in a Test for Markoff Dependency*, Reed B. Dawson, Jr., Dept. of Defense.
17. *A Problem in Combinatory Analysis and its Applications to Probability Theory*, T. B. Narayana, Institut H. Poincare and McGill University, introduced by Harold Hotelling.
18. *The Bayesian Inference Problem in Stochastic Systems*, Max A. Woodbury, George Washington University.

10:30 a.m. Methodology of Studying Motivation. With American Statistical Association

Chairman: Frederick Mosteller, Harvard University

Papers:
1. *Some Statistical Aspects of the Q Technique*, R. R. Bahadur and D. L. Wallace, University of Chicago.
2. *Industrial Mobility of Labor as a Probability Process*, Isadore Blumen, Marvin Kogan, and Philip McCarthy, Cornell University.

Discussion: William Stephenson, Greenwich, Connecticut, Andrew Baggaley, University of Wisconsin, Leo Goodman, University of Chicago

11:00 a.m. Unpublished Mathematical Tables of Interest to Statisticians

Chairman: Dan Teichroew, National Cash Register Company, Dayton, Ohio

Discussion Session

2:00 p.m. Contributed Papers IV

Chairman: Emanuel Parzen, Columbia University

Papers:
1. *Some Nonparametric Generalizations of Multivariate Analysis and Analysis of Variance*, S. N. Roy, University of North Carolina.
2. *"No Interaction" in a Three-Way Table*, Marvin A. Kastenbaum, University of North Carolina.
3. *On Bartlett's Test of Complex Contingency Table Interaction*, Sujit Kumar Mitra, University of North Carolina, (By title).
4. *A Theorem in Minimum Chi Square*, Sujit Kumar Mitra, University of North Carolina, (By title).
5. *Sequential Estimation from a Finite Population*, Herbert David and Ingram Olkin, University of Chicago.
6. *Further Remarks on Measures of Association for Cross-Classifications*, Leo A. Goodman, University of Chicago, and William H. Kruskal, Universities of California and Chicago.

7. *Uniformly Consistent Sequences of Multiple-Decision Rules*, WILLIAM HALL, University of North Carolina, (By title).
8. *Some Hypergeometric Series Distributions Occurring in Birth-and-Death Processes at Equilibrium (Preliminary Report)*, WILLIAM HALL, University of North Carolina, (By title).
9. *Some General Aspects of Stochastic Approximations*, TOSIO KITAGAWA, Iowa State College.
10. *The Analysis of Incidence Rates under Multiple Classifications of the Population (Preliminary Report)*, WYMAN RICHARDSON, University of North Carolina.
11. *Estimation of Percentiles by Order Statistics*, A. E. SARHAN, University of North Carolina, (By title).
12. *On Renewal Theory, Counter Problems, and Quasi-Poisson Processes*, WALTER L. SMITH, University of North Carolina, (By title).
13. *On the Construction of Significance Tests on the Circle and the Sphere*, G. S. WATSON, The Australian National University, Canberra, A.C.T., Australia, (By title).
14. *Estimation of Individual Variations in an Unreplicated Two-Way Classification (Preliminary Report)*, THOMAS S. RUSSELL and RALPH A. BRADLEY, Virginia Polytechnic Institute, (By title).
15. *Empirical Bayes Estimation (Preliminary Report)*, M. V. JOHNS, JR. Columbia University.

A. BIRNBAUM
Associate Secretary

MINUTES OF THE ANNUAL BUSINESS MEETING, 1955

A business meeting was called to order at 6:15 p.m., December 28, 1955 in the Music Room of the Biltmore Hotel, New York City by President Henry Scheffé. Approximately 96 members were present. John Tukey reported on the activities of the Institute Committee on Mathematical Tables. An ad hoc committee to acquaint statisticians with the use of high speed calculators for statistical analysis was announced. R. L. Anderson is chairman of this committee. The Treasurer's report was presented and approved. The Secretary's report was presented and approved. The tellers were instructed to accept ballots from members who had not returned them by mail. The Editor's report was presented and accepted The President presented his report and turned over the chair to the new president, David Blackwell. President Blackwell thanked the outgoing president for his work for the Institute during the past year. A discussion of the advisability of holding Institute of Mathematical Statistics meetings in hotels was conducted as new business. The President was instructed to appoint a committee or to take such other action as he deems suitable to improve physical facilities if meetings are held in hotels.

The tellers announced the election of the following:

President Elect

Alexander M. Mood

Members of IMS Council for term 1955–1958

R. C. Bose	Oscar Kempthorne
Churchill Eisenhart	W. J. Youden

The meeting was adjourned at 7:15 p.m.

G. E. Nicholson, Jr.,
Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1955

The Institute of Mathematical Statistics has by tradition made the last task of the retiring President a review of the year's activity by our society. In 1955 we held in the United States two national and two regional meetings. Besides the present national meeting in New York another was held at Ann Arbor in August and September. An Eastern regional meeting was held at Chapel Hill in April, and a Western regional meeting at Berkeley in July during the time of the Third Berkeley Symposium on Mathematical Statistics and Probability. Because of the Ann Arbor meeting no Central regional meeting was held, but the Central Regional Program Committee worked on plans for a meeting in Chicago in April, 1956. The Program Committees for these meetings are listed with other committees in the appendix to this report; their chairmen, who bore so much of the responsibility, were Herbert Solomon for this meeting, Carl F. Kossack for the Ann Arbor meeting, G. E. Noether for the Chapel Hill meeting, and Herman Rubin for the Berkeley meeting; Theodore A. Bancroft was chairman of the Central Regional Program Committee. Our meetings and the *Annals of Mathematical Statistics* are enriched by Special Invited Papers; the chairman of the Committee that selects authors for these was Herman Chernoff. A new service to our profession may be offered at future meetings in the form of an employment register; the question has been studied by a committee under the chairmanship of Gerald J. Lieberman.

A new policy on meetings was adopted by the Council in September. A Committee to Explore the Desirability of Changing the Time of Winter meetings had been set up last year under the chairmanship of C. C. Craig in response to dissatisfaction expressed by some of our members. On the basis of a sample survey of our North American members the committee judged there was very strong support among our members in all parts of North America for all four of the following propositions: (1) There should be only one national meeting a year. (2) This should alternate between the meeting places of national meetings of the American Mathematical Society and the American Statistical Association. (3) There should be a strengthened program of regional meetings. (4) The preferred

time of meeting is in the first half of September. The Council adopted a tentative policy for the next three years embracing these four propositions.

Since abstracts heretofore have been published in connection with meetings, the new policy implying fewer and more irregularly timed meetings raised questions about separating the publication of abstracts from meetings. This problem has been considered by a committee under the chairmanship of Howard Levene. Their recommendations have been adopted by the Council and include publication of abstracts in every issue of the *Annals*, soon after their receipt.

This year our society organized its first Summer Statistical Institute. It was on the topic "Statistical inference in stochastic processes," was also held in Berkeley during the time of the Third Symposium, and was financed by the National Science Foundation. The idea of the Summer Institutes is to bring together a number of researchers on some specialty for an extended time, to interact with each other to the advancement of the specialty. The benefits hoped for are long range, there being no publications by the Summer Institute as such. David Blackwell was chairman of the Organizing Committee for the 1955 Summer Institute. We hope to run another in 1957.

Discussion of the 1955 Volume of the *Annals of Mathematical Statistics* and its outlook for 1956 I have left to the report by our Editor, T. E. Harris.

Another publishing activity of which we can be very proud is the preparation of the Wald Memorial Volume. This collection of papers by Abraham Wald was published under our sponsorship by the McGraw-Hill Book Company. Chairman of the Editorial Committee for the book was T. W. Anderson.

A further publishing activity has been under consideration by the Committee on Activities and Development under the chairmanship of T. W. Anderson: The University of Chicago is contemplating publication of a series of monographs on mathematical statistics, and our committee has studied possible forms of cooperation by the Institute.

You have heard about our growth in membership during the year and about our financial state in the reports by our Secretary, George E. Nicholson, Jr., and our Treasurer, Albert H. Bowker. Our Individual Memberships Committee under the chairmanship of Eugene W. Pike has been successfully active, campaigning this year especially for graduate students. We gathered no new Institutional Memberships; I suggest every member of the Institute ask himself whether or not he is in a good position to promote one of these, and then follow his conscience.

The former Advisory Committee on Statistical Computations was rejuvenated this year and renamed the IMS Committee on Mathematical Tables. This large committee has been very active under the chairmanship of John W. Tukey in exploring plans for new kinds of service to statisticians. The problems concern the kinds of tables needed, who might compute them, and forms of publication; you will continue to hear about them through this committee. An offshoot will be a new committee on fast machines.

Of especial interest to our members in government service and to those training statisticians is the work of the Committee on Professional Standards of Statis-

ticians in Government Service under the chairmanship of Bradford K. Kimball. The committee is drafting a proposed letter to state personnel boards from the committee, suggesting schedules of duties that can be performed by statisticians at various levels, and corresponding schedules of qualifications.

A Committee to Study the Possibilities of Closer Cooperation with Other Societies under the chairmanship of Samuel S. Wilks has made various specific suggestions, including some about more joint meetings, which will come before the new Council.

The two members who did the most for our society in the last three years, both in volume and importance of their work, retired from office last summer. They are Kenneth J. Arnold, our ex-Secretary-Treasurer, and Erich L. Lehmann, our ex-Editor. Their achievements were memorialized by resolutions passed at the Ann Arbor meeting, which are printed in this issue of the *Annals*. I might add that Arnold is in effect being replaced by three people, our Secretary, George E. Nicholson, Jr., our Treasurer, Albert H. Bowker, and our Program Coordinator, Leo Katz. Beginning this year the Council separated the position of Program Coordinator from that of Program Chairman of the Annual Meeting, with the intention of somewhat lightening the Secretary's burden: The Program Coordinator acts as a kind of secretary for meetings in general. The Council also decided to make the Program Coordinator an Associate Secretary, thus including him in the Council. Our new Editor, T. E. Harris, was the happy choice of a committee under the chairmanship of Samuel S. Wilks, which canvassed the possibilities, personal and institutional, for the Editorship. We are grateful to Michigan State University, the University of California, the University of North Carolina, Stanford University, and The RAND Corporation for encouraging Arnold, Lehmann, Nicholson, Bowker, and Harris to accept these offices, for providing space, and for other forms of assistance.

The Committee on Fellows, under the chairmanship of William G. Madow, in addition to its usual work prepared a report to the Council on standards for election to fellowship. I am pleased to announce that on nomination by this committee the Council has elected the following members to be Fellows:

K. J. Arnold
G. E. P. Box
Gustav Elfving
M. H. Quenouille
Murray Rosenblatt
Herbert Solomon

As the next Nominating Committee I have appointed the following members, who have accepted:

Frederick Mosteller, Chairman
Jerome Cornfield
Harald Cramér
David G. Kendall
Brockway McMillan

I take this opportunity to express my thanks, personally and in behalf of our society, to all those who have faithfully served the Institute: the other officers, the Council members and committee members, our representatives, and the referees for the *Annals*. The referees are listed in the report of the Editor, the others in the following appendix.

HENRY SCHEFFÉ
President

December 28, 1955

Appendix. Committees of the Institute, 1955

I. The Council and Committees of the Council

(a) *Elected members of the Council*

Term expires 1955
W. G. Cochran
Churchill Eisenhart
Henry Scheffé
J. W. Tukey

Term expires 1956
T. W. Anderson, Jr.
Joseph Berkson
Z. W. Birnbaum
David Blackwell
W. G. Madow

Term expires 1957
R. L. Anderson
Leo Goodman
P. G. Hoel
L. J. Savage
Herbert Solomon

(b) *Executive Committee*

President: Henry Scheffé
President-Elect: David Blackwell
Secretary: K. J. Arnold (term expired June 30, 1955)
George E. Nicholson, Jr., (term began July 1, 1955)
Treasurer: K. J. Arnold (term expired March 31, 1955)
Albert H. Bowker (term began April 1, 1955)
Editor: E. L. Lehmann (terminated July 31, 1955)
T. E. Harris (term began August 1, 1955)

(c) *Committee on Fellows*

Term expires 1955
W. G. Madow, *Chairman*
Edward Paulson

Term expires 1956
David Blackwell
Howard Levene

Term expires 1957
M. S. Bartlett
L. J. Savage

(d) *Associate Secretaries*

Evelyn Fix
W. H. Kruskal
Lionel Weiss

(e) *Associate Treasurer*

E. S. Pearson

(f) *Associate Editors*

Z. W. Birnbaum (term began August 1, 1955)
David Blackwell (term expired July 31, 1955)
Herman Chernoff (term began August 1, 1955)
H. E. Daniels (term expired January 1, 1956)
W. J. Dixon (term began August 1, 1955)

J. L. Hodges, Jr. (term expired July 31, 1955)
J. M. Hammersley (term began January 1, 1956)
Wassily Hoeffding (term expired July 31, 1955)
W. G. Madow (term expired July 31, 1955)
L. J. Savage (term began August 1, 1955)
J. Wolfowitz

II. Editorial Committee
(a) T. E. Harris, *Editor*
(b) *Associate Editors*, listed immediately above
(c) *Cooperating Members*

Z. W. Birnbaum	D. A. Darling	G. E. Noether
R. C. Bose	J. L. Doob	M. Peisakoff
G. E. P. Box	T. E. Harris	H. E. Robbins
Herman Chernoff	Paul G. Hoel	L. J. Savage
Kai Lai Chung	J. Kiefer	Charles M. Stein
D. R. Cox	William H. Kruskal	Lionel Weiss
J. F. Daly	Solomon Kullback	Max A.Woodbury

Appointed Committees

III. Program Committees
(a) December Meeting—New York City
Herbert Solomon, *Chairman*
Milton V. Johns, Jr., Asst. Sec.
Kenneth J. Arrow
Donald A. S. Fraser
Oscar Kempthorne
Melvin P. Peisakoff
S. N. Roy

(b) September Meeting—Ann Arbor
Carl F. Kossack, *Chairman*
W. H. Kruskal, Assoc. Sec.
P. S. Dwyer, Asst. Sec.
Henry B. Mann
Horace W. Norton
Paul R. Rider
Murray Rosenblatt

(c) Eastern Region
G. E. Noether, *Chairman*
Lionel Weiss, Assoc. Sec.
Wassily Hoeffding, Asst. Sec.
Robert E. Bechhofer
Ralph A. Bradley
Glenn L. Burrows
J. Edward Jackson

(d) Central Region
Theodore A. Bancroft, *Chairman*
Allen T. Craig
Donald A. Darling
Henry Teicher

(e) Western Region
Herman Rubin, *Chairman*
Evelyn Fix, Assoc. Sec.
Charles H. Kraft, Asst. Sec.
Z. W. Birnbaum
W. J. Dixon
Theodore E. Harris
Stanley W. Nash

(f) Program Coordinator: Leo Katz, (ex officio member of all Program Committees)

(g) Special Invited Paper Committee
Herman Chernoff, *Chairman*
Theodore E. Harris
Wassily Hoeffding
Frederick Mosteller
Max A. Woodbury
E. L. Lehmann (ex officio)
G. E. Nicholson, Jr. (ex officio)

IV. Promotional Committees
(a) Individual Memberships
Eugene W. Pike, *Chairman*
W. D. Baten
Carl A. Bennett
James L. Dolby
Harry M. Hughes

(b) Academic Institutional Memberships
Boyd Harshbarger, *Chairman* — Herbert Robbins
Gerald J. Lieberman — Murray Rosenblatt

(c) Non-Academic Institutional Memberships
Brockway McMillan, *Chairman* — F. W. Dresch
Cuthbert Daniel — Alexander M. Mood

V. Other Committees

(a) Nominating Committee
(Appointed by 1954 President E. G. Olds)
W. J. Dixon, *Chairman* — H. B. Mann
M. A. Girshick — H. Nisselson
M. G. Kendall — A. W. Tucker

(b) Committee on Mathematical Tables
John W. Tukey, *Chairman* — W. J. Dixon
Albert H. Bowker, Vice-Chairman — J. L. Hodges, Jr.
D. Teichroew, Secretary — William Kruskal
R. L. Anderson — H. O. Hartley
Robert E. Bechhofer — J. W. Hopkins
C. I. Bliss — Jack Moshman
Max A. Woodbury

(c) Committee on Activities and Development
T. W. Anderson, *Chairman* — William Kruskal
Joseph Berkson — Samuel S. Wilks
Albert H. Bowker

(d) Committee to Nominate a New Editor
Samuel S. Wilks, *Chairman* — T. W. Anderson
Mina Rees

(e) Committee to Explore the Desirability of Changing Time of Winter Meetings
C. C. Craig, *Chairman* — Carl F. Kossack
R. L. Anderson — R. B. Murphy
J. L. Hodges, Jr.

(f) Committee on Professional Standards of Statisticians in Government Service
B. F. Kimball, *Chairman* — A. S. Householder
Robert W. Burgess — Joseph Lev
Besse B. Day — Herbert Marshall
Churchill Eisenhart — Robert E. Patton
G. M. Harrington — John E. Walsh

(g) Committee to Study the Possibilities of Closer Cooperation with Other Societies
Samuel S. Wilks, *Chairman* — Edwin G. Olds
William G. Cochran

(h) Committee on Committee Procedures
Horace W. Norton, *Chairman* — Morris H. Hansen
K. J. Arnold — G. E. Nicholson, Jr. (ex officio)
Leo A. Goodman — Frederick F. Stephan

(i) Committee to Consider the Feasibility of Keeping an Employment Register at IMS Meetings
Gerald J. Lieberman, *Chairman* — I. R. Savage
David Blackwell

(j) Committee to Study the Possibilities of Arranging a U. S. Visit by Russian Probabilists
David Blackwell, *Chairman* — Jerzy Neyman
J. L. Doob — Herbert Robbins
Eugene Lukacs

(k) Organizing Committee for Summer Statistical Institute in 1955
David Blackwell, *Chairman*
Herbert Robbins
T. E. Harris

(l) Planning Committee for a Summer Statistical Institute in 1956
Herbert Robbins, *Chairman*
Milton Sobel
Jerzy Neyman
John W. Tukey
L. J. Savage

(m) Committee to Consider Desirability of a Summer Statistical Institute in 1956
Frederick Mosteller, *Chairman*
John W. Tukey
Cuthbert Daniel
D. F. Votaw, Jr.
Wassily Hoeffding
William J. Youden
L. J. Savage

(n) Editorial Committee for the Wald Memorial Volume
T. W. Anderson, *Chairman*
E. L. Lehmann
Harald Cramér
Alexander M. Mood
Harold A. Freeman
Charles M. Stein
J. L. Hodges, Jr.

(o) Committee to Consider Policy on Publishing Abstracts
Howard Levene, *Chairman*
Wassily Hoeffding
T. W. Anderson
Jack C. Kiefer

(p) Finance Committee
Mortimer Spiegelman, *Chairman*
A. H. Bowker (ex officio)
K. J. Arnold (ex officio)
John E. Walsh

(q) Advisory Committee on Physical Facilities for Meetings
Z. W. Birnbaum, *Chairman*
George E. Nicholson, Jr.
Leo Katz (Program Coordinator)

(r) Committee on Exchanges
Paul S. Dwyer, *Chairman*
G. E. Nicholson, Jr. (ex officio)
K. J. Arnold (ex officio)
Albert H. Bowker (ex officio)
E. L. Lehmann (ex officio)
T. E. Harris (ex officio)

(s) Committee to Re-examine the Constitution and By-Laws
William G. Cochran, *Chairman*
Tjalling C. Koopmans
T. W. Anderson
Henry Scheffé (ex officio)
Arnold Court
Samuel S. Wilks

(t) Committee to Consider the Format of the *Annals*
Paul G. Hoel, *Chairman*
Alexander M. Mood
George W. Brown

Representatives of the Institute for 1955

To the American Association for the Advancement of Science
Harold Hotelling

To the National Research Council, Division of Mathematics
Samuel S. Wilks

To the Policy Committee for Mathematics
Joseph F. Daly

To the Advisory Committee of American Standards Association concerning ISO/TC 69, Statistical Treatment of Series of Observations
Howard Raiffa

To the Inter-Society Committee on the Mathematical Training of Social Scientists
W. G. Madow, T. W. Anderson

RESOLUTIONS OF THE INSTITUTE

The following resolutions were voted at the Ann Arbor meeting of the Institute of Mathematical Statistics, September 1, 1955.

1) WHEREAS, Kenneth J. Arnold has ably and faithfully served the Institute of Mathematical Statistics in the dual capacity of Secretary-Treasurer from July 1952 through June 1955, and has now retired from this position; and

WHEREAS, in performing the duties of this office Professor Arnold has made great personal sacrifices; and

WHEREAS, besides handling the duties of Treasurer with diligence and foresight, and performing the regular Secretarial duties with competence, tact, and sympathetic consideration for the positions of others, Professor Arnold has made a signal contribution to the efficient operation of the Institute by preparing a *Codification of Actions of the Council* and a *Manual for the Guidance of Officers, Standing Committees, and Representatives* of the Institute: Therefore be it

RESOLVED, that Members of the Institute of Mathematical Statistics at this Sixty-seventh Meeting record their gratitude and appreciation of the high-efficiency, faithfulness, good will, and vision with which Professor Arnold discharged the duties of Secretary-Treasurer; and be it

RESOLVED further, that the President of the Institute of Mathematical Statistics be instructed to forward a copy of this Resolution to John A. Hannah, LL.D., President of the Michigan State University of Agriculture, and Applied Science.

2) RESOLVED, that the membership of the Institute of Mathematical Statistics expresses its thanks to the retiring editor, Erich L. Lehmann, for his able discharge of the difficult and time-consuming duties of his office, and for the continuing development of the *Annals of Mathematical Statistics* under his leadership.

RESOLVED further, that the President of the Institute of Mathematical Statistics be instructed to forward a copy of this Resolution to Dr. Clark Kerr, Chancellor of the University of California at Berkeley.

REPORT OF THE SECRETARY OF THE INSTITUTE FOR 1955

During 1955 the Institute held its sixty-fifth through sixty-eighth meetings. Business meetings were held during the sixty-seventh (seventeenth summer) meeting and the sixty-eighth (eighteenth annual) meeting. The Program Committees are to be congratulated on the excellent programs which have been arranged under the immediate direction of T. A. Bancroft, Carl F. Kossack, G. E. Noether, Herman Rubin and Herbert Solomon with the overall guidance of our

Program Coordinator, Leo Katz. The Assistant Secretaries, P. S. Dwyer, Wassily Hoeffding, Milton V. Johns, Jr., and Charles H. Kraft, are to be congratulated on the physical arrangements, and the Associate Secretaries, Allan Birnbaum, Evelyn Fix, and W. H. Kruskal, on their performance of the duties of the Secretary with respect to these meetings.

In November a supplement to the directory of October 15, 1954, including changes of address and new members as of October 15, 1955, was issued.

G. E. NICHOLSON, JR.,
Secretary

December 28, 1955

REPORT OF THE EDITOR OF THE ANNALS FOR 1955

The 1955 volume of the *Annals* contained 76 papers, of which 16 were notes. This brought the total number of pages, counting miscellaneous material, to 785, a decrease of 41 pages below the previous year. The number of papers submitted in the year ending November 1, 1955, was the same as the number submitted in the preceding twelve-month period. The backlog of accepted manuscripts has grown and now amounts to about one and one-half issues, with some possibility of further growth. It therefore seems desirable, in spite of the rise of printing cost rates in 1955, to use about 900 pages in 1956.

In accordance with a Council decision, the *Annals* will cease to be copyrighted in 1956. Beginning with the March or June issues, names of previous Editors will be listed on the front inside cover, and certain other changes in the inside cover format may be made.

The Editor wishes to thank the previous Editor, E. L. Lehmann, for his generous cooperation, and to acknowledge gratefully the work of David Blackwell, H. E. Daniels, J. L. Hodges, Jr., Wassily Hoeffding, W. G. Madow, and J. Wolfowitz who have continued to act as Associate Editors on manuscripts submitted during the previous term.

Many thanks are due to the Cooperating Members, old and new, and to the following people, (other than Cooperating Members) for very generous refereeing assistance, with apologies to any who are inadvertently omitted: R. L. Anderson, T. W. Anderson, R. R. Bahadur, G. Baxter, R. Bechhofer, A. Birnbaum, J. Blum, D. Chapman, W. Conner, D. R. Cox, G. B. Dantzig, M. Donsker, M. Dwass, C. Eisenhart, B. Epstein, T. Ferguson, Evelyn Fix, D. Fraser, I. J. Good, L. Goodman, H. Kahn, E. Kaplan, O. Kempthorne, D. G. Kendall, L. LeCam, J. Lieblein, E. Lukacs, H. B. Mann, A. Marshall, B. McMillan, E. Paulson, R. L. Plackett, E. S. Pearson, J. Putter, D. Ray, J. Riordan, Joan R. Rosenblatt, M. Rosenblatt, H. Rubin, H. Scheffé, M. Sobel, F. Spitzer, D. Truax, H. Tucker, J. Tukey, A. M. Walker, D. Wishart.

The Editor is especially indebted to Patricia Rice for secretarial work, and to

Berniece Johnson, Dorothy Stewart, and Helena Williams for carrying out the editorial work.

T. E. Harris
Editor

December 28, 1955

PUBLICATIONS RECEIVED

French Bibliographical Digest (Part I: Pure Mathematics), Series 2, No. 14, July, 1955, The Cultural Division of the French Embassy, New York, 128 pp. (Free of charge upon request).

Klein, L. R. and Goldberger, A. S., *An Econometric Model of the United States 1929–1952*, North-Holland Publishing Co., Amsterdam, 1955, xv + 165 pp., $4.50.

Mathematical Models of Human Behavior (Proceedings of a Symposium), 1955, Dunlap and Associates, Inc., Stamford, Connecticut, vii + 103 pp.

Miller, Herman P., *Income of the American People*, John Wiley and Sons, Inc., New York, 1955, 206 pp., $5.50.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

March, 1956

1108 16th St., N.W. Washington 6, D. C. VOL. 51 NO. 273

ASA

THE AMERICAN STATISTICAL ASSOCIATION INVITES AS MEMBERS ALL PERSONS INTERESTED IN:

1. Development of new theory and method
2. Improvement of basic statistical data
3. Application of statistical methods to practical problems.

ESTADISTICA

Journal of the Inter American Statistical Institute

Volume XIII, No. 48–49 Contents September-December 1955

Published quarterly *Annual subscription price* $3.00 (U. S.)

INTER AMERICAN STATISTICAL INSTITUTE

Pan American Union
Washington 6, D.C.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS FOR 1956

President:

David Blackwell, University of California, Berkeley, California

President-Elect:

Alexander M. Mood, General Analysis Corporation, 710 Wilshire Blvd., Santa Monica, California

Secretary:

G. E. Nicholson, University of North Carolina, Chapel Hill, North Carolina

Treasurer:

A. H. Bowker, Stanford University, Stanford, California

Editor:

T. E. Harris, The RAND Corporation, 1700 Main Street, Santa Monica, California

The purpose of the Institute of Mathematical Statistics is to encourage the development, dissemination, and application of mathematical statistics.

Membership dues including a subscription to the ANNALS OF MATHEMATICAL STATISTICS are $10.00 per year for residents of the United States or Canada and $5.00 per year for residents of other countries. There are special rates for students and for some other classes of members. Inquiries regarding membership in the Institute should be sent to the Secretary of the Institute.

MEETINGS OF THE INSTITUTE

TENTATIVE SCHEDULE

EASTERN REGIONAL MEETING—PRINCETON, N. J.
April 20–21, 1956 (With Biometric Society)

CENTRAL REGIONAL MEETING—CHICAGO, ILLINOIS
April 27–28, 1956 (With Biometric Society)

ANNUAL MEETING—SEATTLE, WASHINGTON
August 20–24, 1956 (With American Mathematical Society and others)

MIDWESTERN REGIONAL MEETING—DETROIT, MICHIGAN
September 7–10, 1956 (With American Statistical Association and others)

(See special announcement in this issue concerning submission of Abstracts.)